Genetic Overlap and Causal Mediation Relationship Between Psychiatric and Non-Psychiatric Phenotypes

GENETIC OVERLAP AND CAUSAL MEDIATION RELATIONSHIP BETWEEN PSYCHIATRIC AND NON-PSYCHIATRIC PHENOTYPES

YEN-FENG LIN

A Dissertation Submitted to the Faculty of

The Harvard T.H. Chan School of Public Health

in Partial Fulfillment of the Requirements

for the Degree of *Doctor of Science*

in the Department of *Epidemiology*

Harvard University

Boston, Massachusetts.

March, 2018

GENETIC OVERLAP AND CAUSAL MEDIATION RELATIONSHIP BETWEEN PSYCHIATRIC AND NON-PSYCHIATRIC PHENOTYPES

ABSTRACT

Genome-wide genotyping studies are providing evidence that psychiatric disorders are truly polygenic, that is they have a genetic architecture of many genetic variants. Cross-trait polygenic analysis has been applied to identifying genetic correlations between psychiatric and non-psychiatric phenotypes. However, causal models between shared genetic factors and the genetically-correlated phenotypes are mostly unclear. We used cross-trait polygenic risk score (PRS) association analysis to examine the genetic overlap between two phenotypes. We then performed causal mediation analysis to identify the causal relationship between common genetic variants and two genetically correlated traits. We examined if the effect of polygenic risk on one trait (i.e., the outcome) was mediated by the other trait (i.e., the mediator).

In Chapter 1, we examined the relationship between PRS for psychotic illness or cognitive ability, event-related potential (ERP), and severity of psychotic symptoms. A phenotype of global impairment on multiple ERP measures is associated with positive symptoms of psychosis as well as polygenic influences on educational attainment and, to a lesser extent, schizophrenia. We also observed a positive association between education PRS and positive symptoms that was almost entirely mediated by effects on the globally impaired ERP phenotype.

In Chapter 2, we examined the relationship between PRS for Alzheimer's dementia, vascular pathologies, and late-life cognitive function. Our findings support the hypothesis of a genetic overlap, mostly due to APOE, between vascular pathologies and AD dementia. The polygenic genetic effect on late-life cognition is partially but significantly mediated by cerebral microbleeds, white matter lesion load, and coronary artery calcification.

In Chapter 3, we examined the relationship between PRS for coronary heart disease, psychological attitudes, and liability to coronary heart disease. Our findings suggest a genetic overlap between optimism and CHD in older women of European ancestry. The polygenic genetic effect on CHD is modestly though significantly mediated by optimism.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to express the deepest appreciation to my advisor Dr. Deborah Blacker, for providing me with invaluable opportunities to develop as a scientist, to learn and to advance professionally. Without her patience, insightful mentoring, and persistent help, this dissertation would not have been possible.

I would like to thank Dr. Jordan Smoller, my unofficial advisor, for providing research opportunities, generous support, and guidance. I would also like to thank Dr. Rebecca Betensky and Dr. Alkes Price, for being members of my thesis committee and always offering helpful criticisms about my work and prompted me to think deeper.

I would also like to acknowledge the following collaborators, Mei-Hua Hall (McLean Hospital), Chia-Yen Chen (Massachusetts General Hospital), Lenore Launer (National Institute of Aging), Gudny Eiriksdottir (Icelandic Heart Association; IHA), Albert Smith (IHA), Thor Aspelund (IHA), Vilmundur Gudnason (IHA), and Sylvia Wassertheil-Smoller (Albert Einstein College of Medicine). I am deeply grateful to these people for their kindness in enabling collaborations, guiding my research, and allowing me to use their data.

I greatly appreciate my parents, Chi-Mao Lin and Chao-Chi Chou, for their support and encouragement. I also express thanks to my brother, Yen-Hui Lin, and my sisters-in-law, Erh-Chia Hsieh and Chin-Yu Hsieh, for supporting my work and family. Finally, I want to thank my

wife, Tsai-Yeh Hsieh, and my daughter, Erica Lin, for their companionship and love. I could not have done this without them.

Yen-Feng Lin

Boston, Massachusetts

October, 2017

Chapter 1

Polygenic Pleiotropy and Potential Causal Relationships between Educational Attainment, Neurobiological Profile, and Positive Psychotic Symptoms

Yen-Feng Lin, Chia-Yen Chen, Dost Öngür, Rebecca Betensky, Jordan W. Smoller, Deborah Blacker, and Mei-Hua Hall

ABSTRACT

OBJECTIVE: Event-related portential (ERP) components have been used to assess cognitive functions in patients with psychotic illness. Evidence suggests that among patients with psychosis there is a distinct heritable neurophysiologic phenotypic subtype captured by impairments across a range of ERP measures. In this study, we investigated the genetic basis of this "globally impaired" ERP cluster and its relationship to psychosis and cognitive abilities.

METHODS: We used the K means algorithm applied to six ERP measures to empirically re-derive the globally impaired (n=60) and the non-globally impaired ERP clusters (n=323) in our study sample of cases with schizophrenia (SCZ=136) or bipolar disorder (BPD=121) and healthy controls (n=126). We used published genome-wide association study (GWAS) results for SCZ, BPD, college completion, and childhood intelligence as the discovery datasets to derive polygenic risk scores (PRS) for each of these discovery phenotypes in our study sample and tested associations between each PRS and the globally impaired ERP. We conducted causal mediation analyses to estimate the proportion of each PRS effect on psychotic symptoms (as measured by PANSS positive subscale score) that is mediated through membership in the globally impaired ERP.

RESULTS: Individuals in the globally impaired cluster had significantly higher PANSS-positive scores (β=3.95, P=0.005). The SCZ-PRS was nominally associated with globally impaired ERP (unadjusted P=0.01; R^2=3.07%). We also found a significant positive association between the college-PRS and globally impaired ERP (FDR-corrected P=0.004; R^2=6.15%). The effect of

college-PRS on PANSS-positivity was almost entirely mediated through globally impaired ERP (proportion mediated=97.1%).

CONCLUSION: A phenotype of global impairment on multiple ERP measures is associated with positive symptoms of psychosis as well as polygenic influences on educational attainment and, to a lesser extent, schizophrenia. We also observed a positive association between education PRS and positive symptoms that was almost entirely mediated by effects on the globally impaired ERP phenotype. These results suggest that the globally impaired ERP phenotype may represent some aspects of brain physiology on the causal path between genetic influences on educational attainment and psychotic symptoms

INTRODUCTION

In recent years, traditional psychiatric diagnostic constructs have been increasingly challenged. This is particularly evident in psychotic spectrum disorders such as schizophrenia (SCZ), schizoaffective disorder (SA), and psychotic bipolar disorder (BPD). These disorders overlap substantially in symptoms, neurobiology, cognitive features, treatment response, and liability risk factors (N Craddock, O'Donovan, & Owen, 2005; N. Craddock, O'Donovan, & Owen, 2009; Ivleva et al., 2010; Keshavan, Clementz, Pearlson, Sweeney, & Tamminga, 2013). Moreover, large-scale genetic studies have consistently found overlap in susceptibility across BPD, SA, SCZ, and related phenotypes (Cross-Disorder Group of the Psychiatric Genomics, 2013; Cross-Disorder Group of the Psychiatric Genomics et al., 2013; Ripke et al., 2011; Ruderfer et al., 2014). In line with shared genetic susceptibility, the endophenotype-biomarker literatures on BPD-SCZ indicate differences in degree, rather than differences in kind, across various domains of brain function, both in patients and in their clinically unaffected relatives (Clementz et al., 2016; M. H. Hall, F. Rijsdijk, S. Kalidindi, et al., 2007; M. H. Hall, F. Rijsdijk, M. Picchioni, et al., 2007; Hall et al., 2009; S. K. Hill et al., 2013; Skudlarski et al., 2013; Tamminga et al., 2014; Thaker, 2008). These observations challenge the traditional dichotomous model of SCZ and BPD and support a dimensional approach to understanding how genetic and neurobiological underpinnings cut across diagnostic boundaries.

Auditory event related potential (ERP) components--including P50 sensory gating, N1, P2, and P3--have been extensively investigated in the psychoses and are putative endophenotypes for psychotic spectrum disorders (M.-H. Hall et al., 2007; Hall et al., 2015; Hall et al., 2009; Hall et

al., 2006; O'Connor, Morzorati, Christian, & Li, 1994; O'Donnell, Vohs, Hetrick, Carroll, & Shekhar, 2004; Salisbury, Collins, & McCarley, 2010). Each of these ERP components measures specific aspects of brain function and is reliably quantifiable across diverse clinical and laboratory settings (Owens, Bachman, Glahn, & Bearden, 2016; Tamminga et al., 2014; Turetsky et al., 2007). P50 sensory gating probes inhibitory mechanisms thought to be crucial for protecting the brain from information overload (Freedman et al., 1994). Response to S1 stimulus assesses basic brain functions associated with auditory perception (Javitt, Spencer, Thaker, Winterer, & Hajos, 2008). N1 ERP indexes sensory processing at the level of auditory cortex (Salisbury et al., 2010). P2 and P3 components are associated with higher-order cognitive processes relevant to attention, working memory, and information processing speed (Donchin & Coles, 1988; Polich, 2007) (see Supplementary Information for details on ERPs). In twin analyses, we have shown that ERP phenotypes are heritable and genetically correlated with BPD and SCZ (Hall, 2006, Hall 2007).

We have also identified multivariate clusters of ERP phenotypes that appear to aggregate among individuals with psychotic BPD and SCZ, independent of diagnosis(Hall et al., 2012). In that study, various domains of brain function, ranging from the early pre-attentive stage of information processing to higher complex cognitive processes (including P50 sensory gating, gamma band response, mismatch negativity, and the N1, P2, and P3 ERPs), were included to allow us to empirically derive homogenous subgroups based on these features. One of the clusters was termed “globally impaired” because this group of subjects exhibited functional abnormalities on all of these ERP measures. Such data-driven clustering holds promise for

parsing the neurobiological and genetic heterogeneity of psychotic illness, and the analysis of phenotypes based on these clusters may enhance the power of genetic association analyses (Allison et al., 1998; Marlow et al., 2003; Yang, Williams, & Buu, 2017). Importantly, the neuro-clusters identified in our study resembled the "Biotypes" recently reported by the Bipolar-Schizophrenia Network for Intermediate Phenotypes [BSNIP] consortium (Clementz et al., 2016), even though somewhat different biomarker panels were used in each study. Taken together, these results represent a diagnosis-free approach to integrate information across biomarkers, yielding neurobiologically distinct subgroups, and provide evidence supporting the potential role of neurobiological classification in differentiating individuals with psychotic disorders. The "globally impaired" ERP cluster identified in our prior work was found to be associated with psychotic illness and symptoms across diagnostic boundaries, but its genetic relationship to psychotic illness is unclear.

Findings from well-powered GWAS indicate that psychiatric disorders are highly polygenic, reflecting the influence of thousands of common variants (SNPs) of small effect. Although the individually modest effects of common variants make them uninformative as risk biomarkers, genome-wide polygenic risk scores (PRS), which aggregate the effects of multiple SNPs from GWAS, can capture a substantial liability to disease risk (Krapohl et al., 2016; Rzhetsky, Wajngurt, Park, & Zheng, 2007). The PRS for SCZ and BPD could be used to examine the degree to which multiple risk loci for psychotic illness overlap with those influencing the globally impaired ERP cluster, and such polygenic overlap could provide support for the globally impaired ERP as a putative endophenotype for psychotic illness.

ERP components have been used to assess cognitive functions in patients with psychotic illness. Although cognitive impairment is considered as a core feature of SCZ (Elvevag & Goldberg, 2000) and psychotic BPD (Bortolato, Miskowiak, Kohler, Vieta, & Carvalho, 2015; Daban et al., 2006; Martinez-Aran et al., 2000), the relationship between cognitive performance and SCZ-BPD disorders is complex and controversial. Several epidemiological studies have reported low cognitive ability and poor school performance as risk factors for SCZ and BPD (Agnew-Blais et al., 2015; Glahn, Bearden, Bowden, & Soares, 2006; Jones, Rodgers, Murray, & Marmot, 1994; Loewenstein, Czaja, Bowie, & Harvey, 2012; MacCabe et al., 2008; Osler, Lawlor, & Nordentoft, 2007). However, other studies have found a higher risk of developing psychotic illness among individuals with high levels of cognitive performance and creativity (Andreasen, 1987; Kaufman & Paul, 2014; Kyaga et al., 2013; Kyaga et al., 2011; MacCabe et al., 2010). In addition, recent analysis of cross-trait genetic correlation found a positive genetic correlation between psychotic illness and higher educational attainment (Bulik-Sullivan et al., 2015; Hagenaars et al., 2016; Le Hellard et al., 2016; Okbay et al., 2016), which has been used as a proxy for adolescent and young adult cognitive ability in genetic research (Deary & Johnson, 2010; Rietveld et al., 2013). PRS for both SCZ and BPD were also reported to be highly positively associated with creativity and educational attainment (Power et al., 2015). Therefore, it is worthwhile to further examine the genetic relationship between cognitive ability and ERPs, an electrophysiological index of cognitive functions in psychotic illness.

In the present study, we first used a new sample with a similar panel of ERP phenotypes as those used in our previous study, performing K-means multivariate analyses to derive empirical clusters and see if we could replicate the previously derived neuro-clusters and the association between globally impaired ERP and psychotic symptoms (Hall et al., 2012). We then constructed genome-wide polygenic risk scores (PRS) for psychiatric and cognitive phenotypes, including SCZ, BPD, educational attainment, and childhood intelligence (our discovery phenotypes), to examine the shared genetic components between globally impaired neuro-cluster and each discovery phenotype. Finally, we used a novel approach combining polygenic profiling and causal mediation methods to test the hypothesis that the 'globally impaired' ERP profile represents an intermediate phenotype that mediates genetic influences on the symptomatology of SCZ and BPD. We performed causal mediation analyses to explore whether the 'globally impaired profile' is a mediator between PRS and specific clinical features.

METHODS

Study Sample

The study sample consisted of 258 cases (SCZ =136 and psychotic BPD=122) and 125 healthy controls (prior to genetic quality control procedures). Cases were recruited from McLean Hospital, and healthy controls were recruited through local advertisements. All participants were assessed with the Structured Clinical Interview for DSM Disorders (SCID-I) (M. B. S. First, R.L.; Gibbon M.; Williams, J.B.W., 2002). All participants were of self-reported European ancestry, between 18 and 65 years of age, with no history of neurological disorders, no history of head injury, normal hearing confirmed by audiometric testing, and normal intelligence based on the North American Adult Reading Test (NAART). All cases had no substance abuse (except nicotine) or dependence in the preceding 12 months, did not receive ECT treatment in the preceding 12 months, and were sufficiently stable to participate on an outpatient basis. All controls had no history of psychotic and mood disorders themselves or in a first degree relative, and no substance abuse or dependence in the preceding 12 months. Because of possible genetic overlap between psychosis and mood disorders, the healthy control group included only those free of mood disorders to optimize power to detect genetic risk factors.

This study was approved by the Institutional Review Board at McLean Hospital. Written informed consent was obtained from all participants after fully explaining the aims and procedures of the studies.

Clinical Assessments

All participants completed the SCID-IV diagnostic interview (M. B. First, Spitzer, Gibbon, & Williams, 2002), the Snaith-Hamilton Pleasure Scale (SHPS) (Snaith et al., 1995), and the Mood and Anxiety Symptom Questionnaire (MASQ) (Watson et al., 1995). Demographic (age, sex, years of education, smoking status) and medication information were also obtained from participants. Treatment with antipsychotic medication was quantified in terms of chlorpromazine (CPZ) equivalents (Baldessarini & Davis, 1980). Among the 258 cases, 161 (SCZ n=77, BPD n=84) had data on Positive and Negative Syndrome Scale (PANSS) scores (Kay, Fiszbein, & Opler, 1987), 243 (SCZ n=121, BPD n=122) had data on Young Mania Rating Scale (YMRS) scores (Young, Biggs, Ziegler, & Meyer, 1978) and 138 (SCZ n=66, BPD n=77) had data on Multnomah Community Ability Scale (MCAS) scores (Barker, Barron, McFarland, & Bigelow, 1994). There was no observed association between globally impaired ERP and missing data on each of the rating scales (all $P>0.20$).

Electrophysiological Phenotypic Measures

All participants completed the following tasks: an auditory dual-click paradigm (Adler et al., 1982) and an auditory 'oddball' paradigm (Squires, Squires, & Hillyard, 1975). We applied the same electroencephalogram (EEG) recording and processing procedures as described previously (Hall et al., 2006; Hall et al., 2012). Briefly, EEG was recorded using the BioSemi Active Two system at a digitization rate of 512 Hz, with a bandpass of DC–104 Hz and a Common Mode Sense (CMS) as the reference (PO2 site, parieto-occipital electrode 2) using an 18-channel electrode cap. Blinks and eye movements were monitored through electrodes placed on the left temple and above and below the left eye. The EEG data were re-referenced off-line to the

averaged mastoid. Subjects were not allowed to smoke for a minimum of 40 minutes prior to the recordings. P50 sensory gating and response to S1 stimulus were elicited using the dual-click paradigm. The P50 sensory gating ERP was reported at the Cz site (C = central, z = midline of the head) and calculated as a ratio (S2/S1)×100, where higher ratios reflect more impairment. N1 and P2 amplitude ERPs were elicited by the response to the standard stimuli in the auditory Oddball paradigm and reported at the Cz site, whereas P3 amplitude and latency ERPs were elicited by the response to the target stimuli in the Oddball paradigm reported at the Pz site (P = parietal) (see Supplementary Methods for detail).

K-means Cluster Analyses

As in our previous study (Hall et al., 2012), we included all participants (cases and healthy controls) in the analysis to empirically identify homogeneous subgroups of individuals who share similar neurophysiological profiles, regardless of diagnostic status. Individuals were clustered into 3 distinct sub-groups using the K-means algorithm (Hartigan & Wong, 1979) implemented in JMP (version12.0, SAS Institute Inc.), according to six ERP measures: P50 sensory gating, amplitude of S1 response, N1 amplitude, P2 amplitude, P3 amplitude, and P3 latency. A globally impaired cluster, an intermediate cluster, and a high cognitive functioning cluster were empirically derived (Supplementary Table S1.2). The number of clusters was initially set at 3, based on our previous analysis (Hall et al., 2012). We also applied a V-fold cross-validation method (T. L. Hill, P., 2007) to a range of numbers of clusters (from 2 to 5) and identified 3 as the optimal value of K for K-means.

In our analyses, individuals in the globally impaired cluster were compared to those in the other two clusters (globally impaired [n=60] vs. non-globally impaired [n=323]). We treated the ERP clusters as categories based on our hypothesis that the globally impaired ERP, in particular, may be a useful phenotype for genetic studies.

Genotyping and Quality Control

Genomic DNA from blood samples was extracted by standard procedures at the Massachusetts General Hospital Center for Genomic Medicine. Genotyping was performed at the Broad Institute using the Illumina Infinium OmniExpress array (Illumina Inc.; San Diego, CA, USA). The quality control (QC) procedures have been described elsewhere (Hall et al., 2015). Briefly, we excluded 9 individuals with discordant sex information, missing genotype rate >5% or heterozygosity rate >3 SD, shared IBD >0.125, or non-European ancestry based on principal component analyses. We removed ~45,000 SNPs on the X or Y chromosome, MAF<0.05, call rate <98%, and $P < 1 \times 10E\text{-}6$ for deviation from Hardy-Weinberg equilibrium. The QC steps were carried out with PLINK (S. Purcell et al., 2007) and resulted in a total of 374 subjects with genotype data on 664,907 autosomal SNPs.

We then performed genotype imputation, using the phased haplotypes from the 1,000 Genomes Project dataset as the reference panel. Prephasing and imputation was done with SHAPEIT and IMPUTE2 (Delaneau, Marchini, & Zagury, 2012; Howie, Marchini, & Stephens, 2011). The imputation was performed with the default parameters of the software. The final imputed dataset consisted of 9.7 million autosomal SNPs.

Statistical Analyses

Phenotypic Association Analyses

T-tests, chi-square tests, or multivariable linear regression analyses were used (STATA version12; Stata Corp., College Station, TX) to compare the demographic and clinical characteristics between the globally impaired ERP group and the non-globally impaired ERP group.

PRS Association Analyses

We used GWAS summary statistics for SCZ (Schizophrenia Working Group of the Psychiatric Genomics, 2014) and BPD (Psychiatric, 2011) from the Psychiatric Genomics Consortium (PGC), educational attainment (college completion) (Rietveld et al., 2013) from the Social Science Genetic Association Consortium (SSGAC), and childhood intelligence (Benyamin et al., 2014) from the Childhood Intelligence Consortium (CHIC) as the discovery datasets to derive genome-wide polygenic risk scores (PRS) (S. M. Purcell et al., 2009) for each of the above discovery phenotypes in the study sample. The SCZ discovery sample consisted of 46 non-overlapping case-control samples (33,356 cases and 43,724 controls) and 3 family-based samples (1,396 parent affected-offspring trios). The BPD discovery sample included 11 case-control samples (7,481 cases and 9,250 controls). The college completion discovery sample were combined from 42 GWAS samples (22,475 college and 78,594 non-college), and 95.8% of the individuals were older than 30 years. The childhood intelligence discovery sample consisted of six cohorts

with a total of 12,411 children aged 6 to 18 years. All subjects in the discovery samples were of European ancestry.

To account for only independent association signals from these discovery GWAS, we applied a LD clumping procedure to each discovery dataset, in which we retained the SNP with smallest P-value in each 250 kb window and removed all those in LD (r2> 0.1) with this SNP. We also excluded the Major Histocompatibility Complex (MHC) region between 26 and 33Mb on chromosome 6 when calculating the PRSs, because of the complex haplotype and LD structure in this region. For each discovery phenotype, we used five different association P-value thresholds (PTs)-- 0.001, 0.01, 0.05, 0.1 and 0.5-- to select index SNPs from the clumped independent SNPs for calculating the PRSs. For each individual, we calculated the PRS for each discovery phenotype by summing the risk allele counts of the index SNPs, weighted by the log of their association odds ratios (for SCZ, BPD, and college completion) or the beta coefficients (for childhood intelligence) estimated from the discovery GWAS results.

We used PRSice v1.23 (Euesden, Lewis, & O'Reilly, 2015) to calculate the PRSs and test the association between each PRS and the globally impaired ERP group. Associations were tested using logistic regression models including the top 3 principal components (PCs) of ancestry from the EIGENSTRAT analysis (Price et al., 2006) as covariates. We adjusted for the first 3 PCs because the 4th PC offers very little increase (< 2%) in the total explained variance. Wald test P-values and Nagelkerke's R^2s are reported. We performed the above PRS association analyses on the entire study sample and then repeated the same analyses on the case-only subsample. We

used POLYGENESCORE software in R (Dudbridge, 2013) to calculate statistical power for the association between each PRS and the globally impaired ERP (see Supplementary Methods and Tables S3a-S3d.)

Causal Mediation Analyses

Relationship between PRS, Globally Impaired ERP, and PANSS-Positive Score

For each discovery phenotype that gave evidence of PRS association with globally impaired ERP, we selected the PRS with a P-value threshold that showed the strongest association, and examined its relationship with globally impaired ERP and PANSS positive score in our study sample. We performed regression-based causal mediation analyses to examine whether globally impaired ERP might play a crucial mediating role in the polygenic effect on psychotic symptoms.

In these analyses we estimated the direct effect of each associated PRS (highest vs. lowest quartile) on the PANSS positive score and the indirect effect mediated by globally impaired ERP (binary, globally impaired vs. non-globally impaired), adjusting for the top 3 PCs of ancestry, age, sex, daily chlorpromazine (CPZ) equivalent dose of antipsychotics, and current smoking status at the time of EEG recording, which were potential exposure-mediator, exposure-outcome, or mediator-outcome confounders. The proportion mediated was obtained by dividing the estimated indirect effect by the estimated total effect, as an index of the degree of mediation. This method is based on the counterfactual framework for causal inference (Robins & Greenland, 1992), which is an extension of traditional regression-based mediation

approaches (Baron & Kenny, 1986), allowing binary mediators and outcomes as well as exposure-mediator interactions (T. J. VanderWeele, 2016).

Relationship between PRS, Diagnosis, and Globally Impaired ERP

Because PRS for any of the discovery phenotypes may be associated with the diagnosis of psychotic illness (Cross-Disorder Group of the Psychiatric Genomics, 2013; Okbay et al., 2016), it is possible that the observed relationship between a PRS and globally impaired ERP is a secondary consequence of the PRS effect on psychotic illness. To understand whether the effect of any associated PRS on globally impaired ERP is mediated through "case vs. control status" (i.e., presence vs. absence of psychotic illness) or through one specific major mental illness (SCZ vs. BPD among cases), we also performed mediation analyses to understand the relationships between PRS, diagnosis, and Globally impaired ERP (see Supplementary Methods).

Sensitivity Analyses

Finally, we conducted sensitivity analyses to evaluate the robustness of the above mediation analyses to unmeasured confounding (see Supplementary Methods). All mediation analyses were performed using the PARAMED module in STATA (Emsley & Liu, 2013; Valeri & Vanderweele, 2013). We used bootstrap procedures with 200 replications to compute a 95% bias-corrected bootstrap confidence interval (95% BCCI) for the direct and indirect effects.

RESULTS

Phenotypic associations with globally impaired ERP

The demographic and clinical characteristics of the globally impaired ERP and the non-globally impaired ERP are presented in Table 1.1. In the analysis of all participants, the globally impaired cluster consisted of primarily SCZ or BPD cases (91.7%, which included 48.3% of SCZ cases, 43.3% of BPD cases, vs. 8.3% of controls). The small difference between the proportion of the two disorder groups classified as either globally impaired was not significant. Individuals in the globally impaired cluster were significantly older (P=0.007) and were more likely to be current smokers (P=0.005).

In the analysis restricted to cases only, there was no significant difference in age or other demographic variables between the two ERP clusters. However, SCZ/BPD cases in the globally impaired cluster had significantly higher PANSS-positive scores than those in the non-globally impaired cluster (mean [SD]: 19.88 [7.49] vs. 16.13 [6.98]; P=0.007), and these differences persisted after adjusting for age, sex, daily chlorpromazine equivalent dose of antipsychotics, and smoking status at the time of EEG recording (multivariable linear regression: β=3.95, P=0.005).

Supplementary Table S1.1 presents demographic and clinical information for the study sample by diagnostic group.

PRS Association Analyses

Results of PRS associations between the globally impaired ERP cluster and SCZ-PRS, BPD-PRS, college-PRS, and childhood intelligence-PRS including all subjects are presented in Figure 1.1a and Table 1.2. Results restricted to cases only are presented in Figure 1.1b and Table 1.3. In the full sample analyses, the SCZ-PRS with a P-value threshold of 0.001 (SCZ-$PRS_{PT=0.001}$) was significantly positively associated with risk of globally impaired ERP (unadjusted P=0.01; R^2=3.07%). This association approached significance (FDR-corrected P=0.06) even after correcting for multiple testing by the false discovery rate (FDR) q-value method (J. D. Storey, 2002; J. D. T. Storey, J.E.; Siegmund D., 2004). In the analyses restricted to cases only, results were not significant but were in the same direction (unadjusted P=0.09, R^2=1.76%; FDR-corrected P=0.17). For the BPD-PRS, no significant associations were found with the globally impaired ERP cluster in either the whole sample or the case-only subsample.

In the full sample analyses, we found a significant positive association between the college-PRS and the globally impaired cluster across all five P-value thresholds (Table 1.2, unadjusted P values range from 2.95E-04 to 0.05), such that alleles associated with higher educational attainment were associated with being in the globally impaired cluster. After multiple testing correction, this association remained significant for the college-PRS with a P_T= 0.01 (college-$PRS_{PT=0.01}$, FDR-corrected P=0.004; R^2=6.15%). We also observed a nominally positive association between the childhood intelligence-PRS with P_T= 0.05 and the globally impaired cluster (unadjusted P=0.02, R^2=2.40%; FDR-corrected P=0.08). In the case-only subsample, we again found a significant positive association between the college-$PRS_{PT=0.01}$ and globally impaired ERP membership (Table 1.3, unadjusted P=0.004; FDR-corrected P=0.04; R^2=5.11%)

and a positive association between the childhood intelligence-$PRS_{PT=0.05}$ and globally impaired ERP (unadjusted P=0.01; FDR-corrected P=0.06; R^2=3.70%).

Mediation analyses

Relationship between SCZ-PRS, Globally Impaired ERP, and PANSS-Positive Score

As noted above, the SCZ-PRS with a P-value threshold of 0.001 (SCZ-$PRS_{PT=0.001}$) was nominally associated with the globally impaired cluster, and this association approached significance after correcting for multiple testing. Because of the observed association between globally impaired ERP and PANSS-positive scores among cases, we further examined whether SCZ-$PRS_{PT=0.001}$ is also associated with PANSS-positive score and whether this relationship is mediated by globally impaired ERP (Figure 1.2). The estimated direct and indirect effects betas were 2.68 (95% BCCI: -0.37, 5.52) and 0.27 (95% BCCI: -0.34, 1.23), respectively. The proportion of estimated mediating effect of globally impaired ERP on the total effect of SCZ-PRS on PANSS-positive score was small (9.1%). Adding an exposure-mediator interaction term did not substantially change the effect estimates (direct effect β =2.30 [95% BCCI: -0.82, 5.18]; indirect effect β=0.44 [95% BCCI: -0.63, 1.92]). The minimal effect of including the interaction term suggests that exposure-mediator interaction did not appear to be substantial (T.J. Vanderweele, 2015).

Relationship between college-PRS, Globally Impaired ERP, and PANSS-Positive Score

We found a significant positive association, even after multiple testing correction, between the globally impaired cluster and the college-PRS at the P-value threshold of 0.01. We further examined whether college-$PRS_{PT=0.01}$ is also associated with PANSS-positive score and whether

this relationship is mediated by globally impaired ERP. Results of the analysis with globally impaired ERP cluster as a mediator between college-$PRS_{PT=0.01}$ and PANSS positive symptoms are presented in Figure 1.3. The total effect of the college-$PRS_{PT=0.01}$ on PANSS-positive score was estimated as 0.92 (95% BCCI: -2.62, 5.08). The direct effect was estimated to be β=0.03 [95% BCCI: -3.57, 3.69] and the indirect effect mediated through globally impaired ERP was estimated to be β=0.90 [95% BCCI: 0.11, 2.24] (Figure 1.3). The proportion of mediating effect from college-PRS through globally impaired ERP to PANSS positive was estimated at 97.1%. These results suggest that the effect of the college-$PRS_{PT=0.01}$ on PANSS-positive score was almost entirely mediated through globally impaired ERP. Adding an exposure-mediator interaction term did not substantially change the effect estimates (direct effect β =-0.22 [95% BCCI: -3.97, 3.59]; indirect effect β=1.12 [95% BCCI: 0.06, 3.34]).

Relationship between PRS, Diagnosis, and Globally Impaired ERP

We also examined whether the effect of any associated PRS on globally impaired ERP is mediated through "case vs. control status" or through one specific major mental illness among cases (see Supplementary Results). Nearly one-third (30.9%) of the total effect of SCZ-$PRS_{PT=0.001}$ on globally impaired ERP was mediated by the presence of psychotic illness (Figure S1.1a). Among cases, the proportion of estimated mediating effect of "SCZ vs. BPD" on the total effect of SCZ-$PRS_{PT=0.001}$ on globally impaired ERP was very close to zero (Figure S1.1b). A small proportion (11%) of the total effect of college-$PRS_{PT=0.01}$ on globally impaired ERP was mediated by the presence of psychotic illness (Figure S1.2a). For cases with psychotic illness, the

mediating effect due to having a specific diagnosis of SCZ or BPD was estimated to be zero (Figure S1.2b).

Sensitivity Analyses of Unmeasured Confounding

Sensitivity analyses of unmeasured confounding suggest that even in the presence of strong unmeasured confounding, results of the above mediation analyses would not substantially change (see Supplementary Results and Table S1.4-S1.9). For example, in the mediation analysis with college-PRS as the exposure, globally impaired ERP profile as the mediator, and PANSS positive score as the outcome, existence of unmeasured confounding would likely lead to overestimation of the indirect effect and underestimation of the direct effect. Nonetheless, the estimated indirect effect remained significant after controlling for a strong hypothetical confounder with correlations of 0.3 with both mediator and outcome, and the proportion mediated of 66.5% supported our conclusion that the majority of the effect of college-PRSPT=0.01 on PANSS-positive score was indirect.

DISCUSSION

In the present study, we successfully replicated the clustering of ERP components in an independent sample, including a globally impaired ERP cluster (defined as having abnormalities in all six ERP measures, including P50 sensory gating, amplitude of S1 response, N1 amplitude, P2 amplitude, P3 amplitude, and P3 latency). We also replicated our previous findings that individuals in the globally impaired cluster exhibited greater psychotic symptom severity than individuals in other clusters (Table 1.1) (Hall et al., 2012).

Genetic Overlap between SCZ and globally impaired ERP

Our results demonstrate possible polygenic pleiotropy between SCZ and globally impaired ERP. We found that higher SCZ polygenic risk was marginally associated (unadjusted P-value=0.01, FDR-corrected p = 0.06) with being in the globally impaired ERP cluster. However, there was no observed association between BPD-PRS and the globally impaired ERP cluster.

Globally impaired ERP mediates a small proportion of the effect of SCZ-PRS on PANSS-positive score

Globally impaired ERP was associated with both SCZ-PRS and PANSS positive symptoms score, showing potential to serve as an endophenotype for schizophrenia. Nevertheless, our mediation analysis indicates that only a small proportion (9.1%) of the effect of SCZ-PRS on PANSS-positive score was mediated by globally impaired ERP, suggesting that ERP cluster may not be an ideal intermediate phenotype between SCZ-related genetic variants and positive psychotic symptoms. In addition, we found that the relationship between SCZ-PRS and globally

impaired ERP was significantly mediated by the presence of psychotic illness (i.e., case vs. control status) (see Supplementary Figure S1.1a), implying that the observed association between SCZ-PRS and globally impaired ERP may be only secondary to the effects of SCZ-associated SNPs on the presence of psychotic illness.

Genetic Overlap between higher educational attainment and globally impaired ERP

The evidence for polygenic overlap was strongest for college completion and globally impaired ERP. We found significant positive PRS correlations between greater college-PRS (i.e., greater polygenic loading for higher education) and the globally impaired cluster across all five P-value thresholds (Table 1.2), with the strongest signal at the PRS P-value threshold of 0.01, explaining 6% of the variance in the globally impaired ERP in the full sample (n=383). A similar pattern of genetic overlap was also observed between greater childhood intelligence-PRS and being in the globally impaired cluster. These results were unexpected, as cognitive impairment is common among patients with SCZ and BPD and epidemiological studies have indicated that poor school performance and low cognitive ability are risk factors for SCZ and BPD (Agnew-Blais et al., 2015; Glahn et al., 2006; MacCabe et al., 2008). However, our results are compatible with findings for BPD from the Swedish National School Register of over 900,000 individuals showing that those with excellent school performance had a nearly fourfold increased risk of later BPD compared with those with average grades (MacCabe et al., 2010). Our results are also consistent with recent findings examining genetic overlap between psychiatric diseases and cognitive ability. Studies employing an LD score regression approach to estimate cross-trait genetic correlations found positive genetic correlations between BPD/SCZ risk and educational attainment (Bulik-

Sullivan et al., 2015; Okbay et al., 2016). One possible explanation for the LD score regression results is "case ascertainment bias," such that patients from more educated families were more likely to participate in research. However, in our study, we avoided such case ascertainment bias by using an objective physiological phenotype, which was not phenotypically associated with years of education, and found a significant positive genetic correlation between this psychosis-related trait and higher educational attainment. Further research is needed to replicate and explain the counterintuitive genetic correlation between higher educational attainment and globally impaired ERP.

Globally impaired ERP mediates the effect of college-PRS on PANSS-positive score

We found that the effect of the college-$PRS_{PT=0.01}$ on PANSS-positive score was almost entirely mediated through globally impaired ERP membership (Figure 1.3). It has been suggested that a dimensional classification of psychopathology among patients with SCZ and BPD can better reflect the underlying genetic variation (N. Craddock et al., 2009); therefore, PANSS scores have been used in genetic research to identify the genetic underpinning of specific symptom dimensions of psychotic illness (Sengupta et al., 2016). However, the major disadvantage of using specific symptom-domain scores (e.g., PANSS scores) as the phenotype is that they are very likely to be influenced by treatment, stage of illness, and other environmental factors. Globally impaired ERP as an intermediate phenotype of positive symptoms may be less likely to be influenced by clinical or environmental factors. The nearly complete mediation of the association between college-PRS and PANSS-positive by globally impaired ERP implies that the globally impaired ERP may represent some aspects of brain physiology linking higher education

associated alleles and positive psychotic symptoms. By contrast, only a small proportion of the relationship between SCZ-PRS and PANSS positive score was mediated by globally impaired ERP. It is possible that SNPs affecting positive symptom severity partially overlap with both SCZ-associated and education-associated SNPs, and globally impaired ERP may capture the component of positive symptoms that is genetically correlated with educational attainment. Thus, globally impaired ERP may help stratify the genetic components of psychotic symptoms.

Limitations

The present study has several limitations. First, the PRS approach assumes a linear additive model and does not consider gene-gene interactions that may contribute to the underlying genetic architecture of the phenotypes of interest. Second, the effect estimates from the mediation analyses might be biased due to violation of the un-measured confounding assumption (T. J. VanderWeele, 2016). However, our sensitivity analyses suggest that even with the existence of a strong unmeasured confounder for the mediator-outcome relationship, the results of mediation analyses remained robust. Third, the use of super controls may lead to overestimate the association between SCZ-PRS and globally impaired ERP, because a certain proportion of this association is mediated by the presence of psychotic illness. However, analyses for college-PRS would not be substantially influenced, and results of case-only analyses are robust. Forth, our analyses were restricted to individuals of European ancestry, thus limiting the generalizability of the findings to other ethnic populations. Future research should include a broader range of ethnic populations. Finally, although the causal mediation relationship identified by a statistical approach may imply a mechanistic causality, the true mechanisms

governing the processes from exposure to outcome can only be understood by considering the sufficient cause model (i.e., the identification of a set of minimal conditions that inevitably produce outcome). To look into the black box of causal mechanisms, closer observations, more detailed and extensive data, and more scientific knowledge will be needed.

CONCLUSION

This is the first study, to our knowledge, to demonstrate a causal link between genetic risk scores, ERP phenotype, and positive psychotic symptoms. The results also support prior evidence that college education, a proxy for adolescent and young adult cognitive ability, is genetically correlated with psychotic illness, and suggest a potential physiological role for the multivariate ERP profile in the genetic link between cognitive ability and psychotic symptoms.

References:

Adler, L. E., Pachtman, E., Franks, R. D., Pecevich, M., Waldo, M. C., & Freedman, R. (1982). Neurophysiological evidence for a defect in neuronal mechanisms involved in sensory gating in schizophrenia. *Biol Psychiatry, 17*(6), 639-654.

Agnew-Blais, J. C., Buka, S. L., Fitzmaurice, G. M., Smoller, J. W., Goldstein, J. M., & Seidman, L. J. (2015). Early Childhood IQ Trajectories in Individuals Later Developing Schizophrenia and Affective Psychoses in the New England Family Studies. *Schizophr Bull, 41*(4), 817-823. doi:10.1093/schbul/sbv027

Allison, D. B., Thiel, B., St Jean, P., Elston, R. C., Infante, M. C., & Schork, N. J. (1998). Multiple phenotype modeling in gene-mapping studies of quantitative traits: power advantages. *Am J Hum Genet, 63*(4), 1190-1201. doi:10.1086/302038

Andreasen, N. C. (1987). Creativity and mental illness: prevalence rates in writers and their first-degree relatives. *Am J Psychiatry, 144*(10), 1288-1292. doi:10.1176/ajp.144.10.1288

Baldessarini, R. J., & Davis, J. M. (1980). What is the best maintenance dose of neuroleptics in schizophrenia? *Psychiatry Res, 3*(2), 115-122.

Barker, S., Barron, N., McFarland, B. H., & Bigelow, D. A. (1994). A community ability scale for chronically mentally ill consumers: Part I. Reliability and validity. *Community Ment Health J, 30*(4), 363-383.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Pers Soc Psychol, 51*(6), 1173-1182.

Benyamin, B., Pourcain, B., Davis, O. S., Davies, G., Hansell, N. K., Brion, M. J., . . . Visscher, P. M. (2014). Childhood intelligence is heritable, highly polygenic and associated with FNBP1L. *Mol Psychiatry, 19*(2), 253-258. doi:10.1038/mp.2012.184

Bortolato, B., Miskowiak, K. W., Kohler, C. A., Vieta, E., & Carvalho, A. F. (2015). Cognitive dysfunction in bipolar disorder and schizophrenia: a systematic review of meta-analyses. *Neuropsychiatr Dis Treat, 11*, 3111-3125. doi:10.2147/NDT.S76700

Bulik-Sullivan, B., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., Loh, P. R., . . . Neale, B. M. (2015). An atlas of genetic correlations across human diseases and traits. *Nat Genet, 47*(11), 1236-1241. doi:10.1038/ng.3406

Clementz, B. A., Sweeney, J. A., Hamm, J. P., Ivleva, E. I., Ethridge, L. E., Pearlson, G. D., . . . Tamminga, C. A. (2016). Identification of Distinct Psychosis Biotypes Using Brain-Based Biomarkers. *Am J Psychiatry, 173*(4), 373-384. doi:10.1176/appi.ajp.2015.14091200

Craddock, N., O'Donovan, M. C., & Owen, M. J. (2005). The genetics of schizophrenia and bipolar disorder: dissecting psychosis. *J Med Genet, 42*(3), 193-204.

Craddock, N., O'Donovan, M. C., & Owen, M. J. (2009). Psychosis genetics: modeling the relationship between schizophrenia, bipolar disorder, and mixed (or "schizoaffective") psychoses. *Schizophr Bull, 35*(3), 482-490. doi:sbp020 [pii]

10.1093/schbul/sbp020

Cross-Disorder Group of the Psychiatric Genomics, C. (2013). Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet, 381*(9875), 1371-1379. doi:10.1016/S0140-6736(12)62129-1

Cross-Disorder Group of the Psychiatric Genomics, C., Lee, S. H., Ripke, S., Neale, B. M., Faraone, S. V., Purcell, S. M., . . . Wray, N. R. (2013). Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet, 45*(9), 984-994. doi:10.1038/ng.2711

Daban, C., Martinez-Aran, A., Torrent, C., Tabares-Seisdedos, R., Balanza-Martinez, V., Salazar-Fraile, J., . . . Vieta, E. (2006). Specificity of cognitive deficits in bipolar disorder versus schizophrenia. A systematic review. *Psychother Psychosom, 75*(2), 72-84. doi:10.1159/000090891

Deary, I. J., & Johnson, W. (2010). Intelligence and education: causal perceptions drive analytic processes and therefore conclusions. *Int J Epidemiol, 39*(5), 1362-1369. doi:10.1093/ije/dyq072

Delaneau, O., Marchini, J., & Zagury, J. F. (2012). A linear complexity phasing method for thousands of genomes. *Nat Methods, 9*(2), 179-181. doi:10.1038/nmeth.1785

Donchin, E., & Coles, M. G. H. (1988). Is the P300 Component a Manifestation of Context Updating. *Behavioral and Brain Sciences, 11*(3), 357-374.

Dudbridge, F. (2013). Power and predictive accuracy of polygenic risk scores. *PLoS Genet, 9*(3), e1003348. doi:10.1371/journal.pgen.1003348

Elvevag, B., & Goldberg, T. E. (2000). Cognitive impairment in schizophrenia is the core of the disorder. *Crit Rev Neurobiol, 14*(1), 1-21.

Emsley, R., & Liu, H. (2013). PARAMED: Stata module to perform causal mediation analysis using parametric regression models: Boston College Department of Economics. Retrieved from https://ideas.repec.org/c/boc/bocode/s457581.html

Euesden, J., Lewis, C. M., & O'Reilly, P. F. (2015). PRSice: Polygenic Risk Score software. *Bioinformatics, 31*(9), 1466-1468. doi:10.1093/bioinformatics/btu848

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. (2002). Structured clinical interview for DSM-IV-TR Axis I disorders, research version. In. New York: Biometrics Research, New York State Psychiatric Institute.

Freedman, R., Adler, L. E., Bickford, P., Byerley, W., Coon, H., Cullum, C. M., . . . et al. (1994). Schizophrenia and nicotinic receptors. *Harv Rev Psychiatry, 2*(4), 179-192.

Glahn, D. C., Bearden, C. E., Bowden, C. L., & Soares, J. C. (2006). Reduced educational attainment in bipolar disorder. *J Affect Disord, 92*(2-3), 309-312. doi:10.1016/j.jad.2006.01.025

Hagenaars, S. P., Harris, S. E., Davies, G., Hill, W. D., Liewald, D. C., Ritchie, S. J., . . . Deary, I. J. (2016). Shared genetic aetiology between cognitive functions and physical and mental health in UK Biobank (N=112 151) and 24 GWAS consortia. *Mol Psychiatry, 21*(11), 1624-1632. doi:10.1038/mp.2015.225

Hall, M.-H., Rijsdijk, F. V., Kalidindi, S., Schulze, K., Kravariti, E., Kane, F., . . . Murray, R. (2007). Genetic Overlap Between Bipolar Illness and Event-Related Potentials. *Psychol Med,, 37*(5), 667-678. doi:10.1017/S003329170600972X

Hall, M. H., Chen, C. Y., Cohen, B. M., Spencer, K. M., Levy, D. L., Ongur, D., & Smoller, J. W. (2015). Genomewide association analyses of electrophysiological endophenotypes for schizophrenia and psychotic bipolar disorders: a preliminary report. *Am J Med Genet B Neuropsychiatr Genet, 168B*(3), 151-161. doi:10.1002/ajmg.b.32298

Hall, M. H., Rijsdijk, F., Kalidindi, S., Schulze, K., Kravariti, E., Kane, F., . . . Murray, R. M. (2007). Genetic overlap between bipolar illness and event-related potentials. *Psychol Med, 37*(5), 667-678. doi:10.1017/S003329170600972X

Hall, M. H., Rijsdijk, F., Picchioni, M., Schulze, K., Ettinger, U., Toulopoulou, T., . . . Sham, P. (2007). Substantial shared genetic influences on schizophrenia and event-related potentials. *Am J Psychiatry, 164*(5), 804-812. doi:10.1176/ajp.2007.164.5.804

Hall, M. H., Schulze, K., Rijsdijk, F., Kalidindi, S., McDonald, C., Bramon, E., . . . Sham, P. (2009). Are auditory P300 and duration MMN heritable and putative endophenotypes of psychotic bipolar disorder? A Maudsley Bipolar Twin and Family Study. *Psychol Med, 39*(8), 1277-1287. doi:S0033291709005261 [pii]

10.1017/S0033291709005261

Hall, M. H., Schulze, K., Rijsdijk, F., Picchioni, M., Ettinger, U., Bramon, E., . . . Sham, P. (2006). Heritability and Reliability of P300, P50 and Duration Mismatch Negativity. *Behav Genet, 36*(6), 845-857. doi:10.1007/s10519-006-9091-6

Hall, M. H., Smoller, J. W., Cook, N. R., Schulze, K., Hyoun Lee, P., Taylor, G., . . . Levy, D. L. (2012). Patterns of deficits in brain function in bipolar disorder and schizophrenia: a cluster analytic study. *Psychiatry Res, 200*(2-3), 272-280. doi:10.1016/j.psychres.2012.07.052

Hartigan, J. A., & Wong, M. A. (1979). A K-means clusteringalgorithm. *Applied Statistics, 28*, 100–108.

Hill, S. K., Reilly, J. L., Keefe, R. S., Gold, J. M., Bishop, J. R., Gershon, E. S., . . . Sweeney, J. A. (2013). Neuropsychological Impairments in Schizophrenia and Psychotic Bipolar Disorder: Findings from the Bipolar and Schizophrenia Network on Intermediate Phenotypes (B-SNIP) Study. *Am J Psychiatry*. doi:10.1176/appi.ajp.2013.12101298

Hill, T. L., P. (2007). *STATISTICS: Methods and Applications.* Tulsa, OK.: StatSoft.

Howie, B., Marchini, J., & Stephens, M. (2011). Genotype imputation with thousands of genomes. *G3 (Bethesda), 1*(6), 457-470. doi:10.1534/g3.111.001198

Ivleva, E. I., Morris, D. W., Moates, A. F., Suppes, T., Thaker, G. K., & Tamminga, C. A. (2010). Genetics and intermediate phenotypes of the schizophrenia--bipolar disorder boundary. *Neurosci Biobehav Rev, 34*(6), 897-921. doi:10.1016/j.neubiorev.2009.11.022

Javitt, D. C., Spencer, K. M., Thaker, G. K., Winterer, G., & Hajos, M. (2008). Neurophysiological biomarkers for drug development in schizophrenia. *Nat Rev Drug Discov, 7*(1), 68-83. doi:nrd2463 [pii]

10.1038/nrd2463

Jones, P., Rodgers, B., Murray, R., & Marmot, M. (1994). Child development risk factors for adult schizophrenia in the British 1946 birth cohort. *Lancet, 344*(8934), 1398-1402.

Kaufman, S. B., & Paul, E. S. (2014). Creativity and schizophrenia spectrum disorders across the arts and sciences. *Front Psychol, 5*, 1145. doi:10.3389/fpsyg.2014.01145

Kay, S. R., Fiszbein, A., & Opler, L. A. (1987). The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr Bull, 13*(2), 261-276.

Keshavan, M. S., Clementz, B. A., Pearlson, G. D., Sweeney, J. A., & Tamminga, C. A. (2013). Reimagining psychoses: an agnostic approach to diagnosis. *Schizophr Res, 146*(1-3), 10-16. doi:10.1016/j.schres.2013.02.022

Krapohl, E., Euesden, J., Zabaneh, D., Pingault, J. B., Rimfeld, K., von Stumm, S., . . . Plomin, R. (2016). Phenome-wide analysis of genome-wide polygenic scores. *Mol Psychiatry, 21*(9), 1188-1193. doi:10.1038/mp.2015.126

Kyaga, S., Landen, M., Boman, M., Hultman, C. M., Langstrom, N., & Lichtenstein, P. (2013). Mental illness, suicide and creativity: 40-year prospective total population study. *J Psychiatr Res, 47*(1), 83-90. doi:10.1016/j.jpsychires.2012.09.010

Kyaga, S., Lichtenstein, P., Boman, M., Hultman, C., Langstrom, N., & Landen, M. (2011). Creativity and mental disorder: family study of 300,000 people with severe mental disorder. *Br J Psychiatry, 199*(5), 373-379. doi:10.1192/bjp.bp.110.085316

Le Hellard, S., Wang, Y., Witoelar, A., Zuber, V., Bettella, F., Hugdahl, K., . . . Consortium., S. W. G. o. t. P. G. (2016). Identification of Gene Loci That Overlap Between Schizophrenia and Educational Attainment. *Schizophr Bull*.

Loewenstein, D. A., Czaja, S. J., Bowie, C. R., & Harvey, P. D. (2012). Age-associated differences in cognitive performance in older patients with schizophrenia: a comparison with healthy older adults. *Am J Geriatr Psychiatry, 20*(1), 29-40. doi:10.1097/JGP.0b013e31823bc08c

MacCabe, J. H., Lambe, M. P., Cnattingius, S., Sham, P. C., David, A. S., Reichenberg, A., . . . Hultman, C. M. (2010). Excellent school performance at age 16 and risk of adult bipolar disorder: national cohort study. *Br J Psychiatry, 196*(2), 109-115. doi:10.1192/bjp.bp.108.060368

MacCabe, J. H., Lambe, M. P., Cnattingius, S., Torrang, A., Bjork, C., Sham, P. C., . . . Hultman, C. M. (2008). Scholastic achievement at age 16 and risk of schizophrenia and other psychoses: a national cohort study. *Psychol Med, 38*(8), 1133-1140. doi:10.1017/S0033291707002048

Marlow, A. J., Fisher, S. E., Francks, C., MacPhie, I. L., Cherny, S. S., Richardson, A. J., . . . Cardon, L. R. (2003). Use of multivariate linkage analysis for dissection of a complex cognitive trait. *Am J Hum Genet, 72*(3), 561-570.

Martinez-Aran, A., Vieta, E., Colom, F., Reinares, M., Benabarre, A., Gasto, C., & Salamero, M. (2000). Cognitive dysfunctions in bipolar disorder: evidence of neuropsychological disturbances. *Psychother Psychosom, 69*(1), 2-18. doi:12361

O'Connor, S., Morzorati, S., Christian, J. C., & Li, T. K. (1994). Heritable features of the auditory oddball event-related potential: peaks, latencies, morphology and topography. *Electroencephalogr Clin Neurophysiol, 92*(2), 115-125.

O'Donnell, B. F., Vohs, J. L., Hetrick, W. P., Carroll, C. A., & Shekhar, A. (2004). Auditory event-related potential abnormalities in bipolar disorder and schizophrenia. *International Journal of Psychophysiology, 53*(1), 45-55.

Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., . . . Benjamin, D. J. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature, 533*(7604), 539-542. doi:10.1038/nature17671

Osler, M., Lawlor, D. A., & Nordentoft, M. (2007). Cognitive function in childhood and early adulthood and hospital admission for schizophrenia and bipolar disorders in Danish men born in 1953. *Schizophr Res, 92*(1-3), 132-141. doi:10.1016/j.schres.2007.01.009

Owens, E. M., Bachman, P., Glahn, D. C., & Bearden, C. E. (2016). Electrophysiological Endophenotypes for Schizophrenia. *Harv Rev Psychiatry, 24*(2), 129-147. doi:10.1097/HRP.0000000000000110

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clin Neurophysiol, 118*(10), 2128-2148. doi:S1388-2457(07)00189-7 [pii]

10.1016/j.clinph.2007.04.019

Power, R. A., Steinberg, S., Bjornsdottir, G., Rietveld, C. A., Abdellaoui, A., Nivard, M. M., . . . Stefansson, K. (2015). Polygenic risk scores for schizophrenia and bipolar disorder predict creativity. *Nat Neurosci, 18*(7), 953-955. doi:10.1038/nn.4040

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet, 38*(8), 904-909. doi:10.1038/ng1847

Psychiatric, G. C. B. D. W. G. (2011). Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. *Nat Genet, 43*(10), 977-983. doi:10.1038/ng.943

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., . . . Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet, 81*(3), 559-575. doi:S0002-9297(07)61352-4 [pii]

10.1086/519795

Purcell, S. M., Wray, N. R., Stone, J. L., Visscher, P. M., O'Donovan, M. C., Sullivan, P. F., & Sklar, P. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature, 460*(7256), 748-752. doi:nature08185 [pii]

10.1038/nature08185

Rietveld, C. A., Medland, S. E., Derringer, J., Yang, J., Esko, T., Martin, N. W., . . . Koellinger, P. D. (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science, 340*(6139), 1467-1471. doi:10.1126/science.1235488

Ripke, S., Sanders, A. R., Kendler, K. S., Levinson, D. F., Sklar, P., Holmans, P. A., . . . Gejman, P. V. (2011). Genome-wide association study identifies five new schizophrenia loci. *Nat Genet, 43*(10), 969-976. doi:10.1038/ng.940

Robins, J. M., & Greenland, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology, 3*(2), 143-155.

Ruderfer, D. M., Fanous, A. H., Ripke, S., McQuillin, A., Amdur, R. L., Schizophrenia Working Group of Psychiatric Genomics, C., . . . Kendler, K. S. (2014). Polygenic dissection of diagnosis and clinical dimensions of bipolar disorder and schizophrenia. *Mol Psychiatry, 19*(9), 1017-1024. doi:10.1038/mp.2013.138

Rzhetsky, A., Wajngurt, D., Park, N., & Zheng, T. (2007). Probing genetic overlap among complex human phenotypes. *Proc Natl Acad Sci U S A, 104*(28), 11694-11699. doi:10.1073/pnas.0704820104

Salisbury, D. F., Collins, K. C., & McCarley, R. W. (2010). Reductions in the N1 and P2 auditory event-related potentials in first-hospitalized and chronic schizophrenia. *Schizophr Bull, 36*(5), 991-1000. doi:10.1093/schbul/sbp003

Schizophrenia Working Group of the Psychiatric Genomics, C. (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature, 511*(7510), 421-427. doi:10.1038/nature13595

Sengupta, S. M., MacDonald, K., Fathalli, F., Yim, A., Lepage, M., Iyer, S., . . . Joober, R. (2016). Polygenic Risk Score associated with specific symptom dimensions in first-episode psychosis. *Schizophr Res*. doi:10.1016/j.schres.2016.11.039

Skudlarski, P., Schretlen, D. J., Thaker, G. K., Stevens, M. C., Keshavan, M. S., Sweeney, J. A., . . . Pearlson, G. D. (2013). Diffusion tensor imaging white matter endophenotypes in patients with schizophrenia or psychotic bipolar disorder and their relatives. *Am J Psychiatry, 170*(8), 886-898. doi:10.1176/appi.ajp.2013.12111448

Snaith, R. P., Hamilton, M., Morley, S., Humayan, A., Hargreaves, D., & Trigwell, P. (1995). A scale for the assessment of hedonic tone the Snaith-Hamilton Pleasure Scale. *Br J Psychiatry, 167*(1), 99-103.

Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalogr Clin Neurophysiol, 38*(4), 387-401.

Storey, J. D. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society, Series B, 64*(3), 479–498. doi:10.1111/1467-9868.00346

Storey, J. D. T., J.E.; Siegmund D. (2004). Strong control, conservative point estimation, and simultaneous conservative consistency of false discovery rates: A unified approach. *Journal of the Royal Statistical Society, Series B, 66*(1), 187-205.

Tamminga, C. A., Pearlson, G., Keshavan, M., Sweeney, J., Clementz, B., & Thaker, G. (2014). Bipolar and schizophrenia network for intermediate phenotypes: outcomes across the psychosis continuum. *Schizophr Bull, 40 Suppl 2*, S131-137. doi:10.1093/schbul/sbt179

Thaker, G. K. (2008). Neurophysiological endophenotypes across bipolar and schizophrenia psychosis. *Schizophr Bull, 34*(4), 760-773. doi:sbn049 [pii]

10.1093/schbul/sbn049

Turetsky, B. I., Calkins, M. E., Light, G. A., Olincy, A., Radant, A. D., & Swerdlow, N. R. (2007). Neurophysiological endophenotypes of schizophrenia: the viability of selected candidate measures. *Schizophr Bull, 33*(1), 69-94. doi:sbl060 [pii]

10.1093/schbul/sbl060

Valeri, L., & Vanderweele, T. J. (2013). Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. *Psychol Methods, 18*(2), 137-150. doi:10.1037/a0031034

Vanderweele, T. J. (2015). *Explanation in causal inference : methods for mediation and interaction.* . New York: Oxford University Press.

VanderWeele, T. J. (2016). Mediation Analysis: A Practitioner's Guide. *Annu Rev Public Health, 37*, 17-32. doi:10.1146/annurev-publhealth-032315-021402

Watson, D., Weber, K., Assenheimer, J. S., Clark, L. A., Strauss, M. E., & McCormick, R. A. (1995). Testing a tripartite model: I. Evaluating the convergent and discriminant validity of anxiety and depression symptom scales. *J Abnorm Psychol, 104*(1), 3-14.

Yang, J. J., Williams, L. K., & Buu, A. (2017). Identifying Pleiotropic Genes in Genome-Wide Association Studies for Multivariate Phenotypes with Mixed Measurement Scales. *PLoS One, 12*(1), e0169893. doi:10.1371/journal.pone.0169893

Young, R. C., Biggs, J. T., Ziegler, V. E., & Meyer, D. A. (1978). A rating scale for mania: reliability, validity and sensitivity. *Br J Psychiatry, 133*, 429-435.

Table 1.1: Demographic and clinical characteristics of globally impaired and non-globally impaired clusters

Phenotype Characteristics	**All Subjects**		
	Globally Impaired ERP	**Non-globally Impaired ERP**	**P-value**
Diagnosis			X^2=19.11, **P=7.07E-5**
SCZ, *N(%)*	29 (48.3)	107 (33.1)	
BPD, *N(%)*	26 (43.3)	96 (29.7)	
Unaffected, *N(%)*	5 (8.3)	120 (37.2)	
Sex			X^2=0.18, P=0.68
Female, *N(%)*	30 (50.0)	171 (52.9)	
Age (years), *mean(SD)*	43.58 (14.68)	38.41 (13.25)	t-test, **P=0.007**
Education (years), *mean(SD)*	14.58 (2.22)	14.98 (2.27)	t-test, P=0.22
Current Smoker, *N(%)*	24 (42.1)	77 (24.2)	X^2=7.86, **P=0.005**
MASQ Total, *mean(SD)*	121.27 (36.70)	130.53 (37.62)	t-test, P=0.11 MLR, P=0.62
SHPS, *mean(SD)*	1.09 (1.92)	1.84 (2.87)	t-test, **P=0.02** MLR, P=0.29
	Cases with SCZ or BPD		
	Globally Impaired ERP	**Non-globally Impaired ERP**	**P-value**
Diagnosis			X^2<0.0001, P=1.00
SCZ, *N(%)*	29 (52.7)	107 (52.7)	
BPD, *N(%)*	26 (47.3)	96 (47.3)	
Sex			X^2=2.47, P=0.12
Female, *N(%)*	26 (47.3)	120 (59.1)	
Age (years), *mean(SD)*	45.02 (14.39)	41.38 (12.68)	t-test, P=0.07
Education (years), *mean(SD)*	14.38(2.2)	14.62(2.2)	t-test, P=0.50
Current Smoker, *N(%)*	24 (46.2)	69 (34.3)	X^2=2.48, P=0.12
Age of Onset (years), *mean(SD)*	22.35(8.4)	22.87(8.3)	t-test, P=0.70

CPZ Equivalent Dosage (mg), *mean(SD)*	286.06(336.70)	376.62(508.10)	t-test, P=0.24
PANSS Positive Total, *mean(SD)*	19.88(7.49)	16.13(6.98)	t-test, **P=0.007** MLR, **P=0.005**
PANSS Negative Total, *mean(SD)*	13.18(7.46)	12.22(5.64)	t-test, P=0.42 MLR, P=0.31
PANSS General Total, *mean(SD)*	32.82(8.61)	30.28(9.56)	t-test, P=0.17 MLR, P=0.11
MCAS Total, *mean(SD)*	44.54(8.19)	46.83(5.71)	t-test, P=0.09 MLR, P=0.28
YMRS Total, *mean(SD)*	7.81(12.22)	8.64(11.19)	t-test, P=0.64 MLR, P=0.91
MASQ Total, *mean(SD)*	134.93 (36.9)	136.49 (38.6)	t-test, P=0.81 MLR, P=0.83
SHPS, *mean(SD)*	2.02 (2.96)	1.65 (2.3)	t-test, P=0.39 MLR, P=0.31

X^2: Chi-square statistic.

t-test: two-sample t-test for equal means

MLR: multivariable linear regression for the association between clinical assessments and globally impaired ERP, adjusting for (1) age, sex, case-control status, and current smoking status for all subjects; or (2) age, sex, daily chlorpromazine equivalent dose of antipsychotics, and current smoking status for cases with SCZ or BPD.

All bold values are significant at P <0.05

Table 1.2: Polygenic score association analyses between globally impaired ERP and PRS for each discovery phenotype in all subjects (n=383)

Discovery Phenotype (Dataset)	P_T	NSNP	R^2	Unadjusted P-value	FDR- corrected P-value
SCZ (PGC)	0.001	2518	0.0307	**0.01**	0.06
	0.01	7997	0.0108	0.12	0.23
	0.05	19823	0.0076	0.20	0.28
	0.1	29907	0.0026	0.45	0.50
	0.5	76128	5.258E-05	0.91	0.70
BPD (PGC)	0.001	660	0.0005	0.73	0.63
	0.01	3827	0.0012	0.61	0.56
	0.05	13113	1.823E-05	0.95	0.70
	0.1	22162	0.0003	0.80	0.65
	0.5	68772	0.0013	0.60	0.56
College Completion (SSGAC)	0.001	730	0.0177	**0.05**	0.11
	0.01	4151	0.0615	**2.95E-04**	**0.004**
	0.05	13492	0.0291	**0.01**	0.06
	0.1	22246	0.0179	**0.05**	0.11
	0.5	64444	0.0182	**0.05**	0.11
Childhood Intelligence (CHIC)	0.001	314	0.0014	0.58	0.56
	0.01	2227	0.0081	0.18	0.28
	0.05	8597	0.0240	**0.02**	0.08
	0.1	14828	0.0071	0.21	0.28
	0.5	47552	0.0056	0.27	0.33

P_T: the P-value threshold used in the training dataset.
NSNP: different number of independent SNPs included for calculating the PRS, which is determined by the selection of PT.
R^2: Nagelkerke's pseudo R^2, the proportion of variance in globally impaired ERP in our study sample explained by the PRS.
Unadjusted P-value: the P-value of the test for association between the PRS and globally impaired ERP, before multiple testing correction
FDR corrected P-value: the P-value after multiple testing correction by the FDR q-value method.
All bold values are significant at $P < 0.05$

Table 1.3: Polygenic score association analyses between globally impaired ERP and PRS for each discovery phenotype in cases with SCZ or BPD (n=258)

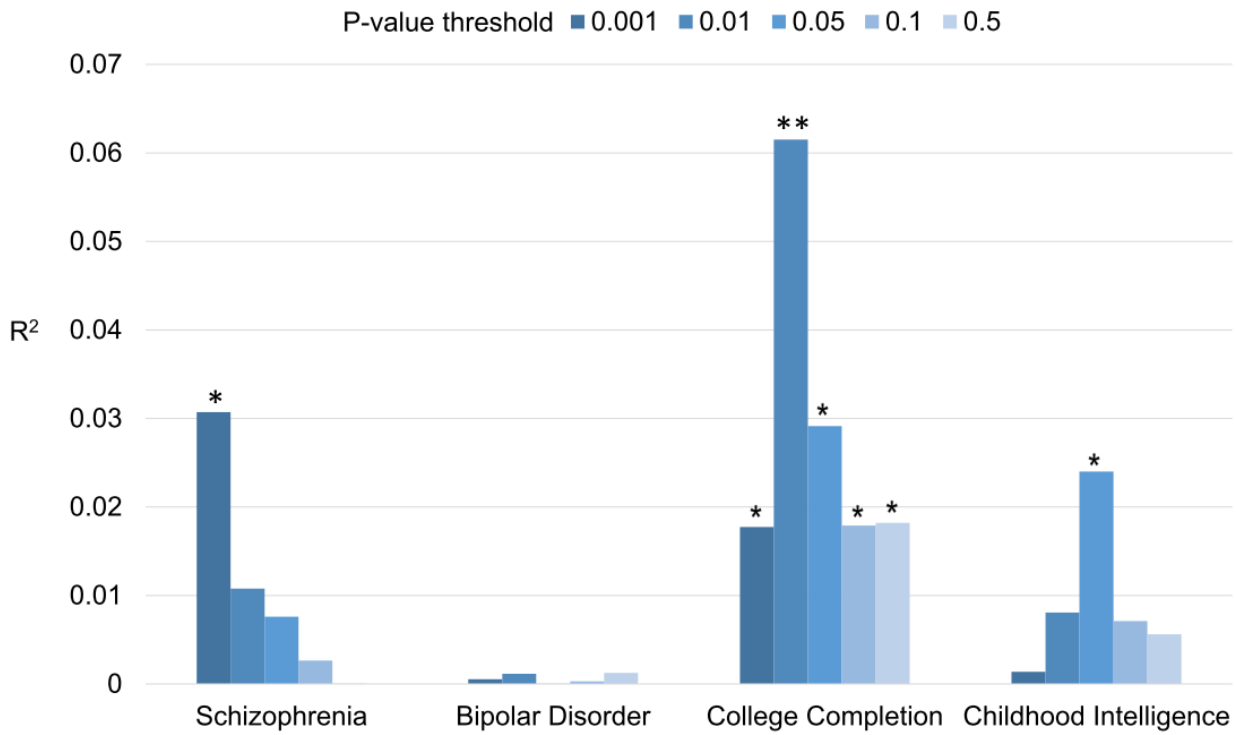
Discovery Phenotype (Dataset)	P_T	NSNP	R^2	Unadjusted P-value	FDR- corrected P-value
SCZ (PGC)	0.001	2518	0.0176	0.09	0.17
	0.01	7997	0.0003	0.84	0.40
	0.05	19823	0.0007	0.73	0.39
	0.1	29907	0.0002	0.87	0.40
	0.5	76128	0.0075	0.27	0.20
BPD (PGC)	0.001	660	0.0001	0.88	0.40
	0.01	3827	0.0077	0.26	0.20
	0.05	13113	0.0041	0.41	0.25
	0.1	22162	0.0053	0.35	0.23
	0.5	68772	0.0010	0.69	0.39
College Completion (SSGAC)	0.001	730	0.0073	0.27	0.20
	0.01	4151	0.0511	**0.004**	**0.04**
	0.05	13492	0.0234	0.05	0.16
	0.1	22246	0.0110	0.18	0.20
	0.5	64444	0.0164	0.10	0.17
Childhood Intelligence (CHIC)	0.001	314	0.0056	0.34	0.23
	0.01	2227	0.0156	0.11	0.17
	0.05	8597	0.0370	**0.01**	0.06
	0.1	14828	0.0109	0.18	0.20
	0.5	47552	0.0102	0.20	0.20

P_T: the P-value threshold used in the training dataset.
NSNP: different number of independent SNPs included for calculating the PRS, which is determined by the selection of P_T.
R^2: Nagelkerke's pseudo R^2, the proportion of variance in globally impaired ERP in patients of our study sample explained by the PRS.
Unadjusted P-value: the P-value of the test for association between the PRS and globally impaired ERP, before multiple testing correction
FDR corrected P-value: the P-value after multiple testing correction by the FDR q-value method.
All bold values are significant at P <0.05

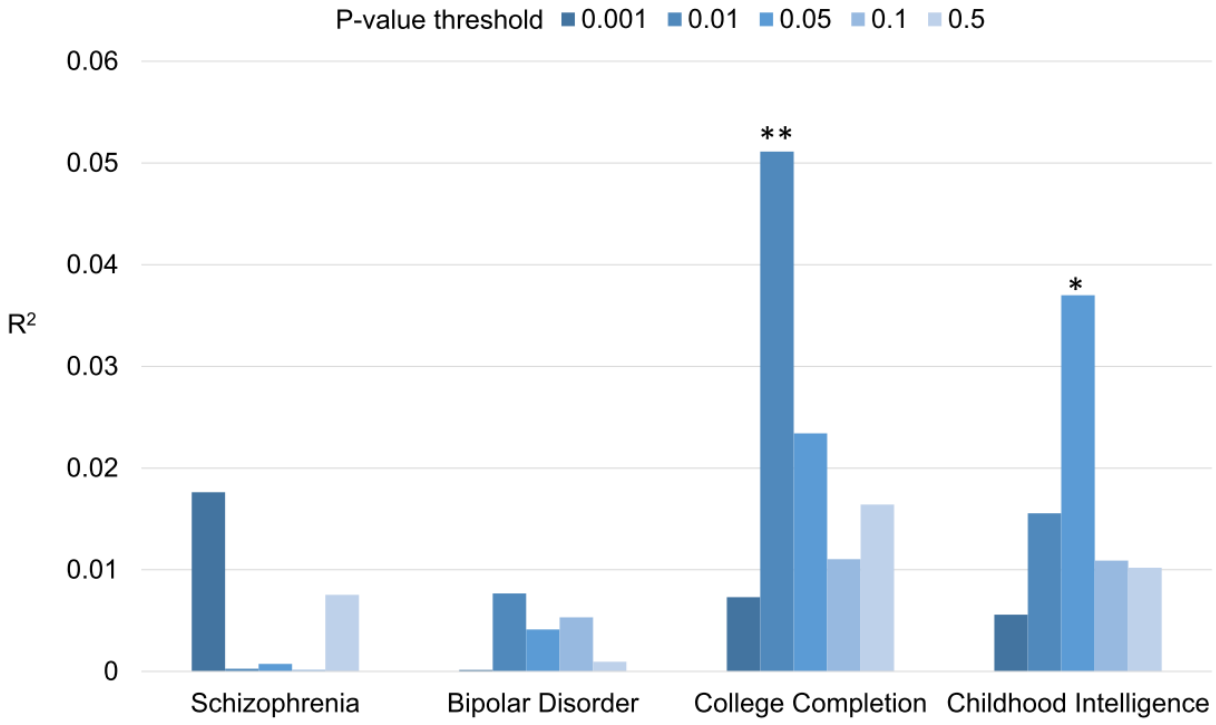
FIGURES

Figure 1.1: Pair-wise polygenic association analyses between globally impaired ERP and PRS for each discovery phenotype

(a) All Subjects



(b) Cases with SCZ or BPD



We derived PRS for schizophrenia, bipolar disorder, college completion, and childhood intelligence from each of the discovery samples with five different P-value thresholds (P_T used to select training set SNPs: 0.001, 0.01, 0.05, 0.1, and 0.5; shown with different colors) and apply them to globally impaired ERP in (A) the entire sample and (B) those affected by SCZ or BPD. Each pair is shown on the x-axis and the proportion of variance explained for globally impaired ERP (estimated via Nagelkerke's pseudo R^2) on the y-axis.
Single asterisk indicates unadjusted P-value <0.05; double asterisk indicates FDR-corrected P-value < 0.05.

Figure 1.2: Causal Relationship between SCZ-PRS, Globally Impaired ERP, and PANSS-Positive Score for patients

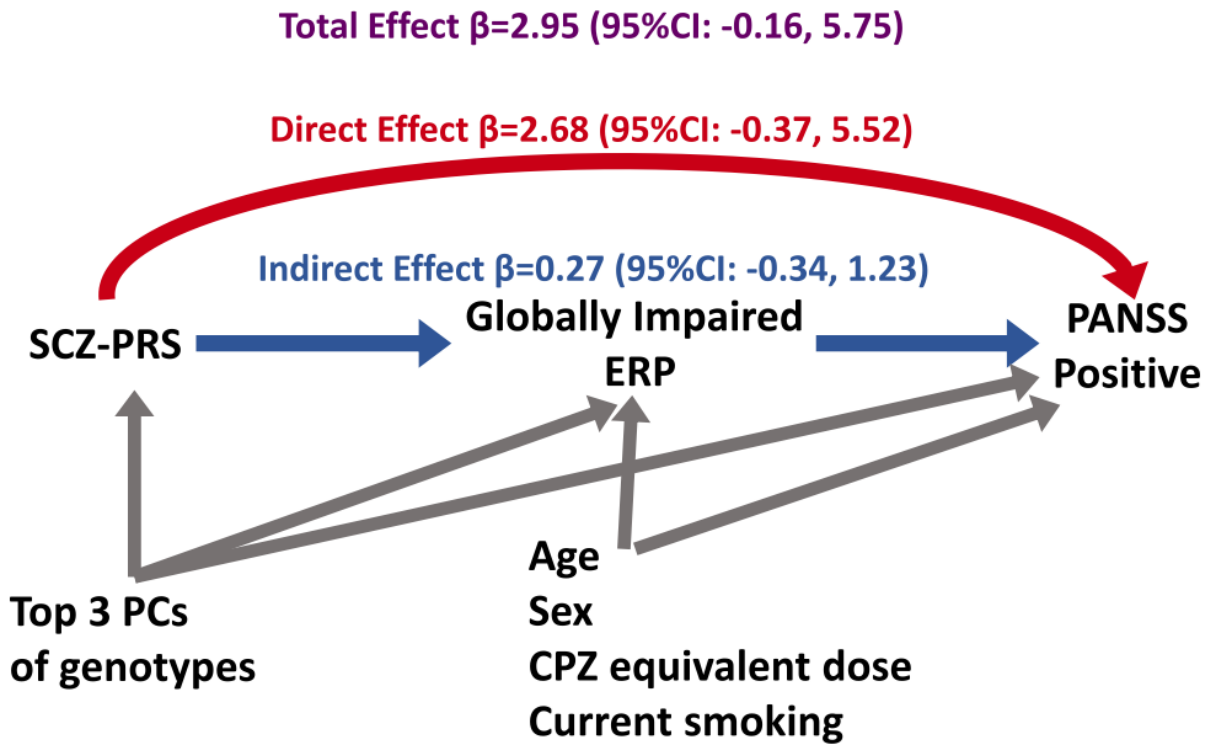
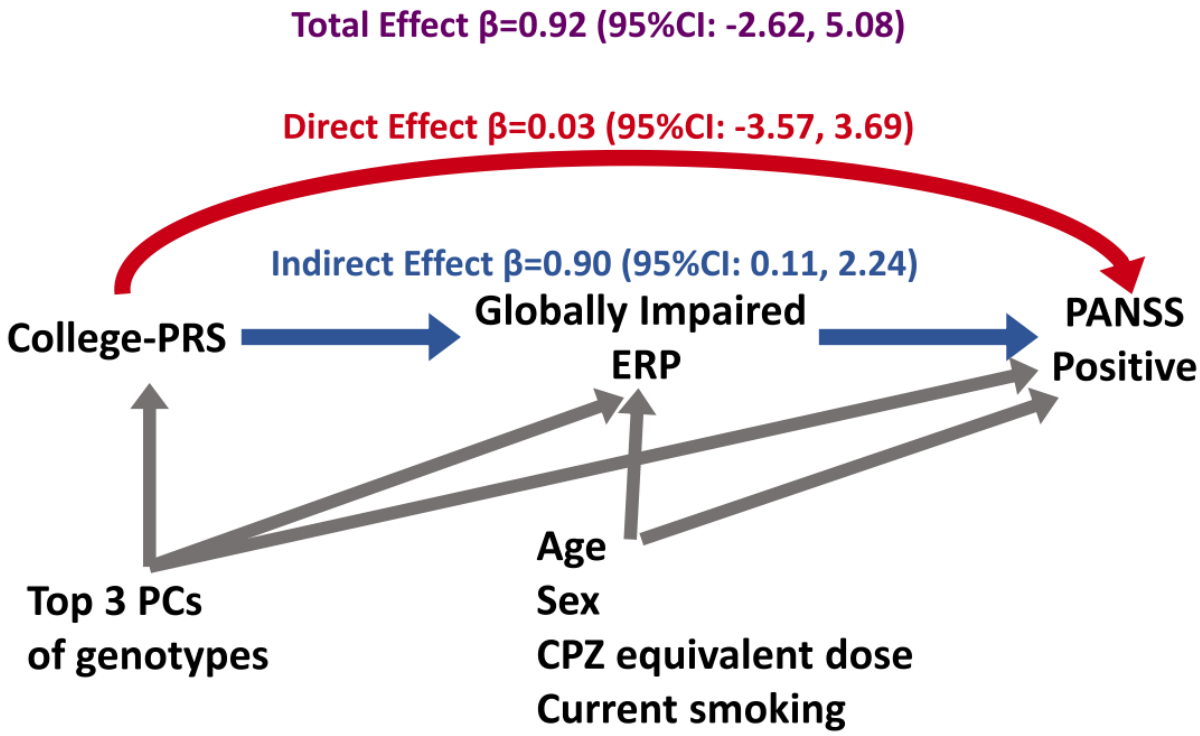


Figure 1.3: Causal Relationship between College-PRS, Globally Impaired ERP, and PANSS-Positive Score for patients



Polygenic Pleiotropy and Potential Causal Relationships between Educational Attainment, Neurobiological Profile, and Positive Psychotic Symptoms

Supplementary Information

SUPPLEMENTARY METHODS

Neurophysiological Recordings and Signal Processing

Event related potential (ERP)

An ERP is a measured brain electrophysiological response to a specific sensory, cognitive, or motor event. ERPs can be reliably measured using electroencephalography (EEG), a procedure that records electrical activity of the brain over time using non-invasive electrodes placed along the scalp.

Dual-Click Paradigm

P50 Sensory Gating. The P50 sensory gating is the brain's suppression of an evoked response to a brief stimulus, usually an auditory click, occurring approximately 50 milliseconds (ms) after receiving the stimulus.

In our study, the P50 sensory gating ERP was elicited using the dual-Click paradigm (160 pairs of identical click stimuli, 5-ms duration; 2-ms rise/fall; 500-ms inter-click interval; 10-s inter-trial interval). Signal processing was performed off-line using NEUROSCAN software (4.3) (Hall et al., 2006; Hall, Taylor, Salisbury, & Levy, 2011).EEG signals were segmented (–100 to 400 ms), filtered (1-Hz high-pass filter), baseline corrected, and artifact rejected if activity exceeding 50 μV between 0 and 75 ms post-stimulus. S1 and S2 waveforms were averaged, digitally filtered (10-Hz high pass), and smoothed. P50 sensory gating ERP are reported at the Cz site and calculated as a ratio (S2/S1)×100. A higher ratio reflects more impairment. For the S1 response, the most prominent peak 40–80 ms post-stimulus was selected as the P50 peak. The preceding negative trough was used to calculate the amplitude. For the S2 response, the positive peak with the latency closest to that of the conditioning P50 peak was selected, and its amplitude was determined as for the S1 wave. P50 sensory gating was calculated as (S2/S1)×100 (Hall et al., 2006; Hall et al., 2011).

Oddball Paradigm

P300 (P3) ERP components. The P3 wave is an endogenous ERP component elicited in the process of decision making. When recorded by electroencephalography (EEG), it surfaces as a positive deflection in voltage with a latency (delay between stimulus and response) of roughly 250 to 500 ms. The signal is typically measured most strongly by the electrodes covering the parietal lobe. It reflects processes involved in stimulus evaluation or categorization. The P3 wave is usually elicited using the oddball paradigm.

In our study, P3 amplitude and latency ERPs were elicited by the auditory Oddball paradigm (400 binaural tones; 50-msec duration, 5 ms rise/fall times; 15% 1500 Hz target tones; 85% 1000 Hz standard tones). All participants had >90% accuracy. Signal processing was performed

off-line using Brain Vision Analyzer software. EEG signals were first re-referenced to linked mastoids and zero phase-shift digital low-pass filtered at 8.5Hz (24 dB/Oct). Eye-blink artifacts were corrected by using the method of Gratton et al. (Gratton, Coles, & Donchin, 1983). The EEG data were segmented into epochs from -100 to 1000 ms relative to stimulus onset and baseline corrected using the 100-ms pre-stimulus interval. Epochs containing artifact >100 µV were removed. Separate average waves for target and standard tones were calculated. P300 amplitude and latency components were measured from the average wave for target tones at the Pz site between 280 and 650 ms (Hall et al., 2009; Salisbury, Shenton, & McCarley, 1999).

N1P2 ERP components. The N1 ERP component is a large, negative-going evoked potential. It peaks in adults between 80 and 120 milliseconds after the onset of a stimulus, and distributed mostly over the fronto-central region. It is elicited by any unpredictable stimulus in the absence of task demands. It is often referred to with the following P2 wave as the "N1-P2" complex. The P2 ERP component is a positive going electrical potential that peaks at about 200 ms (between about 150 and 275 ms) after the onset of external stimulus.
In our study, EEG data were digital low-pass filtered at 20Hz (24 dB/Oct), baseline corrected, eye-blink corrected using (Gratton et al., 1983), and artifact rejected if activity exceeding >100 µV. Peak N1 amplitude was automatically detected as the most negative point from 50 to 200 ms at Cz. Peak P2 amplitude was automatically detected as the most positive point from 150 to 300 ms at Cz (Donchin & Coles, 1988; Polich & Kok, 1995; Salisbury, Collins, & McCarley, 2010).

Power Calculation for the association between PRS and globally impaired

We used POLYGENESCORE software in R (Dudbridge, 2013) to calculate statistical power for the association between each PRS and globally impaired ERP. With known sample sizes (globally impaired/ non-globally impaired: 60/323 in the study sample; case/control: 34,752/46,516 in the SCZ discovery sample, 7,481/9,250 in the BPD discovery sample, and 22,475/ 78,594 in the college education discovery sample; and 12,411 in the childhood intelligence discovery sample), we calculated the power for detecting the PRS association when the genetic correlation between globally impaired and each discovery phenotype is 0.1, 0.3, or 0.5, under the following assumptions:
(1) The prevalence of globally impaired is 0.04, same as the proportion of globally impaired in healthy individuals in our study sample. The prevalence of both SCZ and BPD is 0.01. The prevalence of college completion is 0.22, same as the proportion of college completion in the college education discovery sample.
(2) The SNP-based heritability of SCZ, BPD, college completion, and childhood intelligence is 0.4541, 0.432, 0.0791, and 0.2735, respectively, according to LD score regression analysis results reported on the LD Hub website (http://ldsc.broadinstitute.org/lookup/).
(3) The SNP-based heritability of globally impaired is assumed to be 0.1, 0.3, or 0.5.
(4) The number of independent SNPs in the gene score is assumed to be 1000 or 10000.
(5) The Type-I error of the test for association between the PRS and globally impaired in the target sample is 0.05.

The results of the power analyses are shown in Tables S3a-d.

Supplementary Mediation Analyses

Relationship between PRS, Diagnosis, and Globally Impaired ERP

For each discovery phenotype that gave evidence of PRS association with globally impaired ERP, we selected the PRS with a P-value threshold that showed the highest association, and examined its relationship with the target phenotypes in our study sample. Because PRS for any of the discovery phenotypes may be associated with the diagnosis of psychotic illness (Cross-Disorder Group of the Psychiatric Genomics, 2013; Okbay et al., 2016), it is possible that the observed relationship between a PRS and globally impaired ERP is a secondary consequence of the PRS effect on psychotic illness. To understand whether the effect of any associated PRS on globally impaired ERP is mediated through "case vs. control status" (i.e., presence vs. absence of psychotic illness) or through one specific major mental illness (SCZ vs. BPD among cases), we performed a regression-based causal mediation analysis to decompose the total effect of each PRS on globally impaired into direct and indirect effects, adjusting for potential confounders.

In the first set of mediation analyses, each associated PRS exposure was categorized into quartiles, the potential mediator "case vs. control status" was binary, and the outcome "globally impaired ERP" was treated as a binary variable (globally impaired vs. non-globally impaired). The total effect of each PRS on globally impaired ERP was decomposed into direct and indirect (mediated) effects. These effects were estimated as the odds ratio (OR) for globally impaired comparing the highest quartile to the lowest quartile of the PRS, adjusting for age, sex, and the top 3 PCs of ancestry, which were potential exposure-mediator or exposure-outcome confounders. In the second set of mediation analyses, we performed the same analyses as above in cases, but replaced the mediator variable with diagnosis (SCZ vs. BPD), adjusting for age, sex, and the top 3 PCs of ancestry. We estimated the proportion mediated for each of the above on the log odds scale by dividing the log of the estimated indirect effect OR by the log of the estimated total effect OR (as an index of the degree of mediation).

Sensitivity Analyses of Unmeasured Confounding

The counterfactual-based mediation analysis assumes no unmeasured confounding for the (1) exposure-mediator, (2) exposure-outcome, and (3) mediator-outcome relationships (Vanderweele, 2015). In our mediation analyses with PRS as the exposures, assumptions (1) and (2) were probably plausible, since we had adjusted for the top principal components (PCs) of genotypes to address possible population stratification. However, the assumption of no unmeasured confounding might be less plausible for the (3) mediator-outcome relationship, and the effect estimates would probably be biased.

In order to evaluate the robustness of the mediation analyses to unmeasured confounding between mediator and outcome, we conducted sensitivity analyses to calculate how much direct and indirect effect estimates would be expected to change under different degrees of mediator-outcome confounding. Specifically, given a hypothetical unmeasured confounder of the mediator-outcome relationship, U, with particular correlations with the mediator and the outcome, we would like to know, if we were able to also adjust for U, what the direct and indirect effect estimates would be.

For each mediation analysis, we generated four standard normal variables (i.e., mean=0 and variance=1), with particular correlations with the mediator and the outcome, as hypothetical confounders (Table S1.4-1.9). The first hypothetical confounder has correlation of 0.1 with the mediator and correlation of 0.1 with the outcome. The second hypothetical confounder has correlation of 0.3 with the mediator and correlation of 0.1 with the outcome. The third hypothetical confounder has correlation of 0.1 with the mediator and correlation of 0.3 with the outcome. The fourth hypothetical confounder has correlation of 0.3 with the mediator and correlation of 0.3 with the outcome. We compared the direct and indirect effect estimates before and after adjusting for each of these hypothetical confounders, to assess the potential impact of unmeasured confounding on each mediation analysis.

GWAS of globally impaired ERP

We also performed a preliminary GWAS for globally impairer ERP (globally impaired vs. non-globally impaired). With such a small sample size and lack of replication, we are aware that the GWAS results may not be reliable. We did this just to make the best use of our data and hope to contribute to future research.
We tested each genotyped and imputed SNP for association with globally impaired ERP group in the form of logistic regression assuming an underlying additive model in PLINK (Purcell et al., 2007). We included the top 3 PCs from the EIGENSTRAT analysis (Price et al., 2006) as covariates. We obtained an estimated odds ratio (OR) and a P-value for the association test for each SNP.

SUPPLEMENTARY RESULTS

Supplementary Mediation Analyses

Relationship between SCZ-PRS, Diagnosis, and Globally Impaired ERP

The SCZ-PRS with a P-value threshold of 0.001 (SCZ-PRS$_{PT=0.001}$) was significantly associated with the globally impaired cluster, and this association approached significance after correcting for multiple testing. Thus, as described in Supplementary Methods, we then conducted causal mediation analyses to determine whether the effect of SCZ-PRS$_{PT=0.001}$ on globally impaired ERP was mediated by the presence of SCZ and BPD (together or individually). The results examining whether case vs. control status mediates the relationship between SCZ-PRS$_{PT=0.001}$ and globally impaired ERP are presented in Figure S1.1a. The estimated direct effect OR was 1.76 (95% BCCI: 0.72, 3.80). The indirect effect OR was 1.29 (95% BCCI: 1.12, 1.54), significantly greater than 1. Nearly one-third (30.9%) of the total effect of SCZ-PRS$_{PT=0.001}$ on globally impaired ERP was mediated by the presence of psychotic illness. Adding an exposure-mediator interaction term resulted in a minimal change in the effect estimates (direct effect OR =-1.67 [95% BCCI: 0.77, 3.62]; indirect effect OR=1.36 [95% BCCI: 1.12, 1.65]).

Figure S1.1b presents the results examining in the case only sample whether specific diagnosis (SCZ vs. BPD) mediates the relationship between SCZ-PRS$_{PT=0.001}$ and globally impaired ERP. The estimated direct and indirect effects ORs were 2.28 (95% BCCI: 0.94, 6.08) and 1.00 (95% BCCI: 0.94, 1.10), respectively. The proportion of estimated mediating effect of "SCZ vs. BPD among cases" on the total effect of SCZ-PRS$_{PT=0.001}$ on globally impaired ERP was very close to zero (0.2%).

Relationship between college-PRS, Diagnosis, and Globally Impaired ERP

We found a significant positive association, even after multiple testing correction, between college-PRS (at P_T = 0.01) and the globally impaired cluster. Since patients with SCZ or BPD were more likely to have globally impaired ERP, it is possible that the observed association between the college-PRS and globally impaired ERP could be partly explained by the presence of psychotic illness. We therefore performed two causal mediation analyses to understand whether the effect of college-PRS$_{PT=0.01}$ on globally impaired ERP was mediated by diagnostic status. The results examining whether the effect of college-PRS $_{PT=0.01}$ on globally impaired ERP was mediated by case vs. control status are presented in Figure S1.2a. The estimated direct effect OR was 4.09 (95% BCCI: 1.60, 10.59), while the estimated indirect effect OR mediated by case vs. control status was 1.19 (95% BCCI: 1.00, 1.46) (Figure 1.2a). When an interaction between college-PRS $_{PT=0.01}$ and case vs. control status was included in the regression model, the estimated direct and indirect effect ORs were 4.12 (95% BCCI: 1.66, 10.25) and 1.16 (95% BCCI: 0.98, 1.38), respectively. The minimal effect of including the interaction term suggests that exposure-mediator interaction did not appear to be substantial (Vanderweele, 2015). Overall, then, the effect of college-PRS $_{PT=0.01}$ on globally impaired ERP appeared to be primarily explained by the direct (non-mediated) relationship, whereas the proportion of estimated mediating effect of psychotic illness on the total effect was small (11.1%).

The results examining whether specific diagnosis (SCZ vs. BPD among cases) mediates the relationship between college-$PRS_{PT=0.01}$ and globally impaired ERP are presented in Figure S1.2b. The estimated direct and indirect effects ORs were 3.94 (95% BCCI: 1.47, 10.71) and 1.00 (95% BCCI: 0.93, 1.08), respectively. The mediating effect due to diagnosis was estimated to be zero, indicating that for cases with psychotic illness, the effect of the college-$PRS_{PT=0.01}$ on globally impaired ERP is not mediated by having a diagnosis SCZ or BPD *per se*.

In the full sample, the effect of college-PRS on globally impaired ERP was only modestly (11.1%) mediated by case vs. control status (combining SCZ and BPD cases; Figure S1.2a). While college-$PRS_{PT=0.01}$ was associated with globally impaired ERP, it was not associated with psychotic illness. One possible explanation is that the diagnosis of psychotic illness is heterogeneous, and the ERP phenotype captures the component of psychosis that is correlated with education-associated genes. If we hypothesize that there is some genetic overlap between higher education and psychotic illness, the ERP phenotype may be a better alternative phenotype than traditional diagnosis to detect such genetic overlap. Among cases with psychotic illness, the effect of college-$PRS_{PT=0.01}$ on globally impaired ERP did not appear to be mediated by the specific diagnosis (SCZ vs. BPD) (Figure S1.2b), implying that the ERP phenotype may potentially identify genetically relevant groups independent of the diagnostic boundary between SCZ and BPD.

Sensitivity Analyses of Unmeasured Confounding

SCZ-PRS as the exposure:

Sensitivity analyses of unmeasured confounding found similar results after adjusting for hypothetical confounders in all three mediation analyses with SCZ-PRS as the exposure (Table S1.4-1.6).

In the analysis on patients with SCZ-PRS as the exposure, globally impaired ERP as the mediator, and PANSS positive score as the outcome reported in the main text, the estimated direct and indirect effects odds ratios (95%CI) before adjusting for the hypothetical unmeasured confounder were 2.68 (-0.37, 5.52) and 0.27 (-0.34, 1.23), respectively. The effect estimates after adjusting for each hypothetical confounder U are shown in Table S1.4. Under adjustment of a strong hypothetical confounder with correlations of 0.3 with both mediator and outcome, the estimated direct and indirect effects odds ratios (95%CI) were 2.49 (-0.21, 5.77) and 0.15 (-0.18, 1.04), respectively, which were very close to the unadjusted effect estimates.

In the supplementary analysis with SCZ-PRS as the exposure, case vs. control status as the mediator, and globally impaired ERP profile as the outcome reported in the main text, the estimated direct and indirect effects odds ratios (95%CI) before adjusting for the hypothetical unmeasured confounder were 1.76 (0.72, 3.80) and 1.29 (1.12, 1.54), respectively. The effect estimates after adjusting for each hypothetical confounder U are shown in Table S1.5. Under adjustment of a strong hypothetical confounder with correlations of 0.3 with both mediator and outcome, the estimated direct and indirect effects odds ratios (95%CI) were 1.75 (0.75, 3.99) and 1.24 (1.08, 1.50), respectively, which were very close to the unadjusted effect estimates.

In the supplementary analysis on patients with SCZ-PRS as the exposure, diagnosis (SCZ vs. BPD) as the mediator, and globally impaired ERP profile as the outcome reported in the main text, the estimated direct and indirect effects odds ratios (95%CI) before adjusting for the hypothetical unmeasured confounder were 2.28 (0.94, 6.08) and 1.00 (0.94, 1.10), respectively. The effect estimates after adjusting for each hypothetical confounder U are shown in Table S1.6. Under adjustment of a strong hypothetical confounder with correlations of 0.3 with both mediator and outcome, the estimated direct and indirect effects odds ratios (95%CI) were 2.47 (0.69,7.07) and 0.99 (0.82, 1.09), respectively, which were very close to the unadjusted effect estimates.

College-PRS as the exposure:

In the mediation analysis with college-PRS as the exposure, globally impaired ERP profile as the mediator, and PANSS positive score as the outcome, the estimated direct and indirect effects betas (95%CI) before adjusting for the hypothetical unmeasured confounder were 0.03 (-3.57, 3.69) and 0.90 (0.11, 2.24), respectively. The effect estimates after adjusting for each hypothetical confounder U are shown in Table S1.7. Under adjustment of a strong hypothetical confounder with correlations of 0.3 with both mediator and outcome, the estimated direct and

indirect effects betas (95%CI) were 0.25 (-3.09, 3.83) and 0.49 (0.02, 1.61), respectively. The sensitivity analysis for this mediation analysis indicated that existence of unmeasured confounding would likely lead to overestimation of the indirect effect and underestimation of the direct effect. Nonetheless, the estimated indirect effect remained significant after controlling for a strong hypothetical confounder, and the proportion mediated of 66.5% supported our conclusion that the majority of the effect of college-$PRS_{PT=0.01}$ on PANSS-positive score was indirect.

Sensitivity analyses of unmeasured confounding found similar results after adjusting for hypothetical confounders in the two supplementary mediation analyses with college-PRS as the exposure (Table S1.8 & S1.9).

In the supplementary analysis with college-PRS as the exposure, case vs. control status as the mediator, and globally impaired ERP profile as the outcome reported in the main text, the estimated direct and indirect effects odds ratios (95%CI) before adjusting for the hypothetical unmeasured confounder were 4.09 (1.60, 10.59) and 1.19 (1.00, 1.46), respectively. The effect estimates after adjusting for each hypothetical confounder U are shown in Table S1.8. Under adjustment of a strong hypothetical confounder with correlations of 0.3 with both mediator and outcome, the estimated direct and indirect effects odds ratios (95%CI) were 3.81 (1.40, 10.53) and 1.15 (1.01, 1.41), respectively, which were very close to the unadjusted effect estimates.

In the supplementary analysis on patients with college-PRS as the exposure, diagnosis (SCZ vs. BPD) as the mediator, and globally impaired ERP profile as the outcome reported in the main text, the estimated direct and indirect effects odds ratios (95%CI) before adjusting for the hypothetical unmeasured confounder were 3.94 (1.47, 10.71) and 1.00 (0.93, 1.08), respectively. The effect estimates after adjusting for each hypothetical confounder U are shown in Table S1.9. Under adjustment of a strong hypothetical confounder with correlations of 0.3 with both mediator and outcome, the estimated direct and indirect effects odds ratios (95%CI) were 4.37 (1.31, 13.55) and 1.02 (0.91, 1.24), respectively, which were very close to the unadjusted effect estimates.

GWAS of globally impaired ERP

The GWAS results of globally impaired cluster showed no evidence for genomic inflation (lambda-GC of 0.98, Figure S1.3). Although none of the SNPs reached genome-wide significance (P <5E-08), five independent regions including nine SNPs showed suggestive association levels (p < 1E-05) (Figure S1.4; Table S1.10). Among the suggestive associated SNPs, rs1424104 and rs4888926 are located in the WWOX gene on chromosome 16, rs4792136 and rs73284773 are located in the SHISA6 gene on chromosome 17, and rs1078008 is located in the VIPR1 gene on chromosome 3. A gain-type copy number variation (CNV) affecting the WWOX gene has been found exclusively in patients with SCZ (Rodriguez-Santiago et al., 2010). Although the other two genes have not been reported to be associated with psychotic disorders, both are involved in aspects of brain function.
Again, we are aware that the GWAS is underpowered and the results may not be reliable. Therefore, we did not report them in the main text. Replication with larger sample sizes is required.

SUPPLEMENTARY TABLES

Table S1.1: Socio-demographic characteristics of subject groups

	SCZ Patients	BPD Patients	Healthy Controls
	N=136	N=122	N=125
Age, yrs	44.26(12.36)	39.81(13.59)	33.15(12.54)
Female, N (%)	90(66.2)	56(45.9)	55(44.0)
Education, yrs	14.22(2.12)	14.99(2.30)	15.61(2.17)
Current Smoker, N (%)	51(38.6)	42(34.7)	8(6.6)
Age of Onset	22.94(7.85)	22.54(8.86)	--
CPZ Equivalent Dosage (mg)	516.98(577.33)	214.49(303.28)	--
PANSS Total	62.45(18.79)	57.96(16.49)	--
MCAS Total	45.15(7.15)	47.54(5.17)	--
YMRS Total	6.52(8.55)	10.38(13.42)	--
MASQ Total	141.09(40.88)	131.75(35.09)	101.20(21.86)
SHPS	1.93(2.34)	1.58(2.53)	0.35(0.96)

Note: Values are means (SD) unless otherwise indicated.

Table S1.2: Mean (SD) of ERP measures in each cluster

ERP measures	All Subjects		
	Globally Impaired N=60	**Intermediate N=221**	**High Cognitive N=102**
P50 Sensory Gating	74.79 (37.76)	66.95 (42.40)	37.26 (25.72)
Response to S1	2.25 (1.08)	2.22 (0.88)	4.45 (1.48)
N1 Amplitude	-3.64 (2.70)	-3.89 (2.34)	-6.15 (4.00)
P2 Amplitude	4.80 (3.00)	4.68 (2.81)	9.81 (3.84)
P3 Amplitude	5.92 (0.52)	9.39 (4.26)	14.12 (5.85)
P3 Latency	559.73 (62.43)	377.19 (36.94)	364.15 (34.75)
	Patients Only		
	Globally Impaired N=55	**Intermediate N=162**	**High cognitive N=41**
P50 Sensory Gating	78.55 (37.06)	73.03 (44.77)	49.55 (28.42)
Response to S1	2.20 (1.10)	2.22 (0.91)	4.62 (1.52)
N1 Amplitude	-3.56 (2.76)	-3.67 (2.24)	-5.83 (4.32)
P2 Amplitude	4.95 (3.01)	4.31 (2.80)	9.17 (3.50)
P3 Amplitude	5.66 (4.00)	8.70 (4.07)	12.72 (5.83)
P3 Latency	558.56 (62.98)	380.26 (38.78)	373.06 (35.39)

For P50 sensory gating, a lower value indicates better inhibition. For the response to S1, P2 amplitude, and P3 amplitude measures, a higher value indicates larger responses. For the N1 amplitude, a lower value indicates larger responses. For the P3 latency, a lower value indicates faster processing speed.

Table S1.3a: Results of power calculation for detecting the association between the SCZ-PRS and the globally impaired ERP

The genetic correlation between the discovery and the target traits	Heritability of globally impaired (assumed)	NSNP (assumed)	R^2	P-value	Power
0.1	0.1	1000	9.8E-04	0.26	0.08
		10000	8.7E-04	0.27	0.08
	0.3	1000	0.0030	0.19	0.14
		10000	0.0026	0.20	0.13
	0.5	1000	0.0049	0.13	0.20
		10000	0.0043	0.15	0.18
0.3	0.1	1000	0.0089	0.07	0.33
		10000	0.0078	0.08	0.29
	0.3	1000	0.027	0.005	0.75
		10000	0.023	0.008	0.69
	0.5	1000	0.044	3.7E-04	0.93
		10000	0.039	8.1E-04	0.89
0.5	0.1	1000	0.025	0.007	0.71
		10000	0.022	0.01	0.66
	0.3	1000	0.074	5.1E-06	0.99
		10000	0.065	1.9E-05	0.99
	0.5	1000	0.12	3.0E-09	1.00
		10000	0.11	2.9E-08	1.00

NSNP: Different number of independent SNPs included for calculating the PRS, which is determined by the selection of P-threshold.
R^2: Squared correlation between the PRS and the globally impaired ERP
P-value: Expected p-value of the test for association between the PRS and the globally impaired ERP
Power: The power for detecting the association between the PRS and the globally impaired ERP

Table S1.3b: Results of power calculation for detecting the association between the BPD-PRS and the globally impaired ERP

The genetic correlation between the discovery and the target traits	Heritability of globally impaired (assumed)	NSNP (assumed)	R^2	P-value	Power
0.1	0.1	1000	9.3E-04	0.27	0.08
		10000	5.6E-04	0.28	0.07
	0.3	1000	0.0027	0.19	0.13
		10000	0.0017	0.23	0.10
	0.5	1000	0.0046	0.14	0.19
		10000	0.0028	0.19	0.14
0.3	0.1	1000	0.0084	0.08	0.31
		10000	0.0051	0.13	0.21
	0.3	1000	0.025	0.006	0.72
		10000	0.015	0.03	0.51
	0.5	1000	0.044	2.5E-05	0.98
		10000	0.038	9.7E-05	0.96
0.5	0.1	1000	0.023	0.008	0.69
		10000	0.014	0.03	0.48
	0.3	1000	0.070	9.5E-06	0.99
		10000	0.042	5.0E-04	0.91
	0.5	1000	0.12	9.0E-09	1.00
		10000	0.07	8.3E-06	0.99

NSNP: Different number of independent SNPs included for calculating the PRS, which is determined by the selection of P-threshold.
R^2: Squared correlation between the PRS and the globally impaired ERP
P-value: Expected p-value of the test for association between the PRS and the globally impaired ERP
Power: The power for detecting the association between the PRS and the globally impaired ERP

Table S1.3c: Results of power calculation for detecting the association between the college-PRS and the globally impaired ERP

The genetic correlation between the discovery and the target traits	Heritability of globally impaired (assumed)	NSNP (assumed)	R^2	P-value	Power
0.1	0.1	1000	8.0E-04	0.27	0.07
		10000	2.9E-04	0.30	0.06
	0.3	1000	0.0024	0.20	0.12
		10000	8.7E-04	0.27	0.08
	0.5	1000	0.0040	0.15	0.17
		10000	0.0015	0.24	0.09
0.3	0.1	1000	0.0072	0.09	0.28
		10000	0.0026	0.20	0.13
	0.3	1000	0.022	0.01	0.66
		10000	0.0079	0.08	0.30
	0.5	1000	0.036	0.001	0.87
		10000	0.013	0.04	0.45
0.5	0.1	1000	0.020	0.01	0.63
		10000	0.0073	0.09	0.28
	0.3	1000	0.060	3.7E-05	0.98
		10000	0.022	0.01	0.66
	0.5	1000	0.10	9.5E-08	1.00
		10000	0.036	0.001	0.87

NSNP: Different number of independent SNPs included for calculating the PRS, which is determined by the selection of P-threshold.
R^2: Squared correlation between the PRS and the globally impaired ERP
P-value: Expected p-value of the test for association between the PRS and the globally impaired ERP
Power: The power for detecting the association between the PRS and the globally impaired ERP

Table S1.3d: Results of power calculation for detecting the association between the childhood intelligence-PRS and the globally impaired ERP

The genetic correlation between the discovery and the target traits	Heritability of globally impaired (assumed)	NSNP (assumed)	R^2	P-value	Power
0.1	0.1	1000	7.7E-04	0.25	0.08
		10000	2.5E-04	0.29	0.06
	0.3	1000	0.0023	0.17	0.16
		10000	7.6E-04	0.26	0.08
	0.5	1000	0.0039	0.11	0.23
		10000	0.0013	0.22	0.11
0.3	0.1	1000	0.0070	0.06	0.37
		10000	0.0023	0.17	0.15
	0.3	1000	0.021	0.002	0.81
		10000	0.0068	0.06	0.37
	0.5	1000	0.035	0.0001	0.96
		10000	0.011	0.02	0.56
0.5	0.1	1000	0.019	0.003	0.78
		10000	0.0063	0.06	0.35
	0.3	1000	0.058	7.2E--07	1.00
		10000	0.019	0.004	0.78
	0.5	1000	0.097	9.4E-11	1.00
		10000	0.032	0.0002	0.94

NSNP: Different number of independent SNPs included for calculating the PRS, which is determined by the selection of P-threshold.
R^2: Squared correlation between the PRS and the globally impaired ERP
P-value: Expected p-value of the test for association between the PRS and the globally impaired ERP
Power: The power for detecting the association between the PRS and the globally impaired ERP

Table S1.4: The estimated direct and indirect effect beta (95% CI) of the relationship between the SCZ-PRS, globally impaired ERP profile, and PANSS positive score after adjusting for a hypothetical confounder U

		rUY	
		0.1	**0.3**
rUM	**0.1**	βDE= 2.65 (-0.46, 5.46) βIE= 0.26 (-0.26, 1.35)	βDE= 2.45 (-0.53, 5.39) βIE= 0.23 (-0.28, 1.25)
	0.3	βDE= 2.67 (-0.10, 5.62) βIE= 0.24 (-0.21, 1.29)	βDE= 2.49 (-0.21, 5.77) βIE= 0.15 (-0.18, 1.04)

rUM: the point biserial correlation coefficient between the hypothetical confounder and the mediator

rUY: the Pearson's correlation coefficient between the hypothetical confounder and the outcome

βDE: direct effect beta

βIE: indirect effect beta

The estimated direct and indirect effects betas (95%CI) before adjusting for the hypothetical unmeasured confounder were 2.68 (-0.37, 5.52) and 0.27 (-0.34, 1.23), respectively.

Table S1.5: The estimated direct and indirect effect odds ratio (95% CI) of the relationship between the SCZ-PRS, case-control status, and globally impaired ERP profile after adjusting for a hypothetical confounder U

		rUY	
		0.1	**0.3**
rUM	**0.1**	ORDE= 1.76 (0.73, 3.92) ORIE= 1.29 (1.12, 1.53)	ORDE= 1.74 (0.71, 3.99) ORIE= 1.27 (1.08, 1.54)
	0.3	ORDE= 1.76 (0.74, 3.88) ORIE= 1.31 (1.12, 1.59)	ORDE= 1.75 (0.75, 3.99) ORIE= 1.24 (1.08, 1.50)

rUM: the point biserial correlation coefficient between the hypothetical confounder and the mediator

rUY: the point biserial correlation coefficient between the hypothetical confounder and the outcome

ORDE: direct effect odds ratio

ORIE: indirect effect odds ratio

The estimated direct and indirect effects odds ratios (95%CI) before adjusting for the hypothetical unmeasured confounder were 1.76 (0.72, 3.80) and 1.29 (1.12, 1.54), respectively.

Table S1.6: The estimated direct and indirect effect odds ratio (95% CI) of the relationship between the SCZ-PRS, diagnosis (SCZ vs. BPD), and globally impaired ERP profile after adjusting for a hypothetical confounder U

		rUY	
		0.1	**0.3**
rUM	**0.1**	ORDE= 2.33 (0.80, 5.46) ORIE= 1.00 (0.93, 1.09)	ORDE= 2.46 (0.77, 6.73) ORIE= 1.00 (0.86, 1.05)
	0.3	ORDE= 2.33 (0.87, 6.04) ORIE= 1.00 (0.88, 1.06)	ORDE= 2.47 (0.69,7.07) ORIE= 0.99 (0.82, 1.09)

rUM: the point biserial correlation coefficient between the hypothetical confounder and the mediator

rUY: the point biserial correlation coefficient between the hypothetical confounder and the outcome

ORDE: direct effect odds ratio

ORIE: indirect effect odds ratio

The estimated direct and indirect effects odds ratios (95%CI) before adjusting for the hypothetical unmeasured confounder were 2.28 (0.94, 6.08) and 1.00 (0.94, 1.10), respectively.

Table S1.7: The estimated direct and indirect effect beta (95% CI) of the relationship between the college-PRS, globally impaired ERP profile, and PANSS positive score after adjusting for a hypothetical confounder U

		rUY	
		0.1	**0.3**
rUM	**0.1**	βDE= 0.13 (-3.45, 3.54) βIE= 0.84 (0.18, 2.23)	βDE= 0.27 (-2.91, 4.04) βIE= 0.78 (0.14, 2.16)
	0.3	βDE= 0.09 (-3.56, 3.77) βIE= 0.71 (0.11, 2.15)	βDE= 0.25 (-3.09, 3.83) βIE= 0.49 (0.02, 1.61)

rUM: the point biserial correlation coefficient between the hypothetical confounder and the mediator

rUY: the Pearson's correlation coefficient between the hypothetical confounder and the outcome

βDE: direct effect beta

βIE: indirect effect beta

The estimated direct and indirect effects betas (95%CI) before adjusting for the hypothetical unmeasured confounder were 0.03 (-3.57, 3.69) and 0.90 (0.11, 2.24), respectively.

Table S1.8: The estimated direct and indirect effect odds ratio (95% CI) of the relationship between the college-PRS, case-control status, and globally impaired ERP profile after adjusting for a hypothetical confounder U

		rUY	
		0.1	**0.3**
rUM	**0.1**	ORDE= 4.02 (1.65, 10.35) ORIE= 1.19 (1.02, 1.48)	ORDE= 3.75 (1.35, 10.87) ORIE= 1.17 (1.00, 1.45)
	0.3	ORDE= 4.06 (1.60, 12.03) ORIE= 1.21 (1.02, 1.52)	ORDE= 3.81 (1.40, 10.53) ORIE= 1.15 (1.01, 1.41)

rUM: the point biserial correlation coefficient between the hypothetical confounder and the mediator

rUY: the point biserial correlation coefficient between the hypothetical confounder and the outcome

ORDE: direct effect odds ratio

ORIE: indirect effect odds ratio

The estimated direct and indirect effects odds ratios (95%CI) before adjusting for the hypothetical unmeasured confounder were 4.09 (1.60, 10.59) and 1.19 (1.00, 1.46), respectively.

Table S1.9: The estimated direct and indirect effect odds ratio (95% CI) of the relationship between the college-PRS, diagnosis (SCZ vs. BPD), and globally impaired ERP profile after adjusting for a hypothetical confounder U

		rUY	
		0.1	**0.3**
rUM	**0.1**	ORDE= 4.05 (1.48, 10.51) ORIE= 1.00 (0.93, 1.08)	ORDE= 4.35 (1.38, 13.15) ORIE= 1.00 (0.92, 1.11)
	0.3	ORDE= 4.06 (1.63, 11.53) ORIE= 1.00 (0.93, 1.13)	ORDE= 4.366 (1.308, 13.554) ORIE= 1.02 (0.91, 1.24)

rUM: the point biserial correlation coefficient between the hypothetical confounder and the mediator

rUY: the point biserial correlation coefficient between the hypothetical confounder and the outcome

ORDE: direct effect odds ratio

ORIE: indirect effect odds ratio

The estimated direct and indirect effects odds ratios (95%CI) before adjusting for the hypothetical unmeasured confounder were 3.94 (1.47, 10.71) and 1.00 (0.93, 1.08), respectively.

Table S1.10: Suggestive associated SNPs for Globally Impaired ERP

CHR	SNP	A1	A2	FRQ	OR	SE	P
3	rs1078008	T	C	0.7892	0.3075	0.2534	3.27E-06
6	rs79617003	A	G	0.037	9.7991	0.4788	1.88E-06
8	rs62514812	T	C	0.9226	0.2283	0.3321	8.70E-06
8	chr8_58374581_I	I2	D	0.0595	5.5597	0.3819	7.04E-06
8	rs10081508	T	C	0.0771	4.4419	0.3358	8.96E-06
16	rs4888926	T	C	0.3311	2.7612	0.2289	9.09E-06
16	rs1424104	T	C	0.3234	2.8247	0.23	6.35E-06
17	rs4792136	A	C	0.0752	5.098	0.355	4.48E-06
17	rs73284773	A	G	0.0718	4.8004	0.3447	5.34E-06

SUPPLEMENTARY FIGURES

Figure S1.1a: Causal Relationship between SCZ-PRS, Case vs. Control Status, and Globally Impaired ERP for all subjects

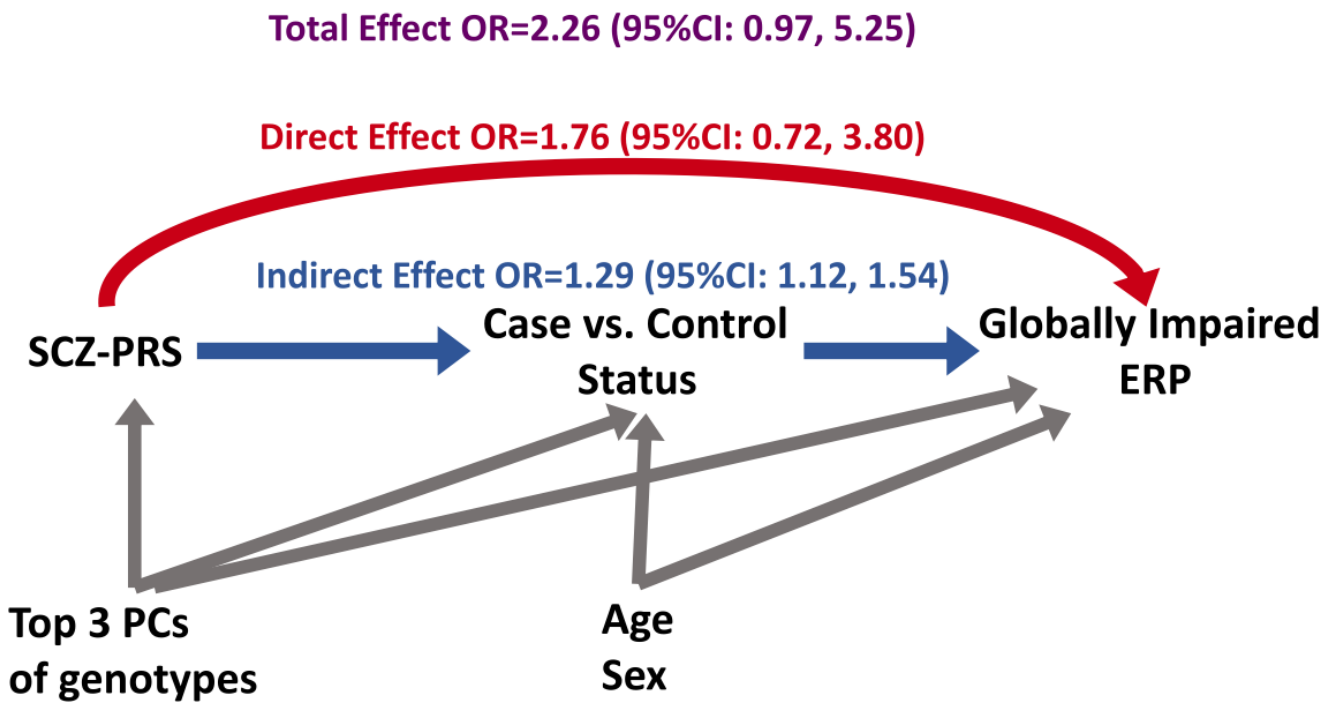


Figure S1.1b: Causal Relationship between SCZ-PRS, SCZ vs. BPD Diagnosis, and Globally Impaired ERP for patients

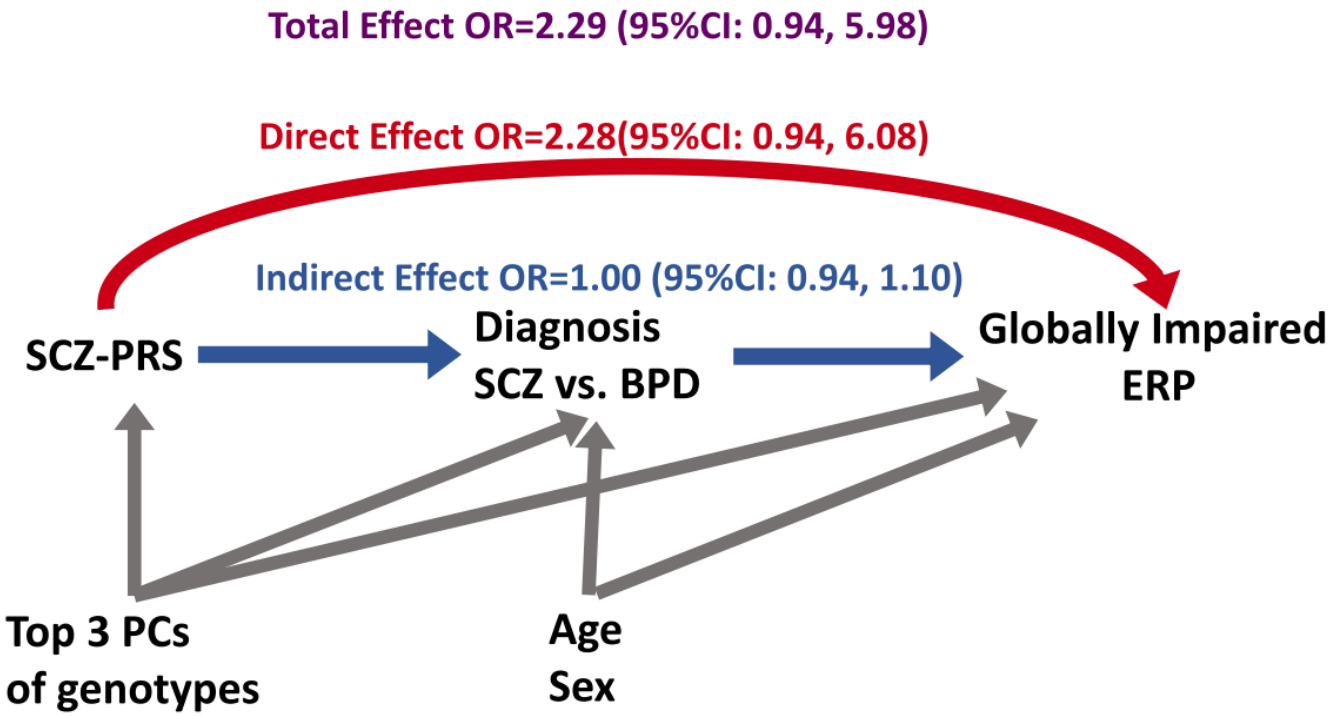


Figure S1.2a: Causal Relationship between College-PRS, Case vs. Control Status, and Globally Impaired ERP for all subjects

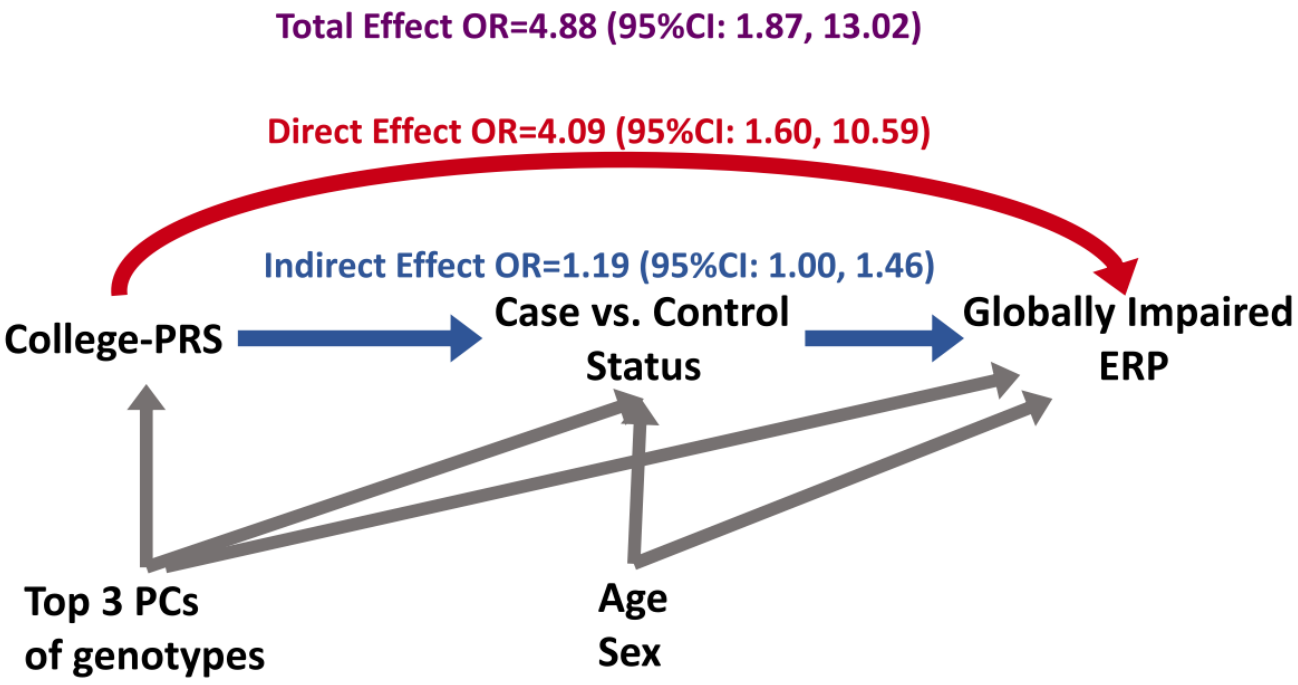


Figure S1.2b. Causal Relationship between College-PRS, SCZ vs. BPD Diagnosis, and Globally Impaired ERP for patients

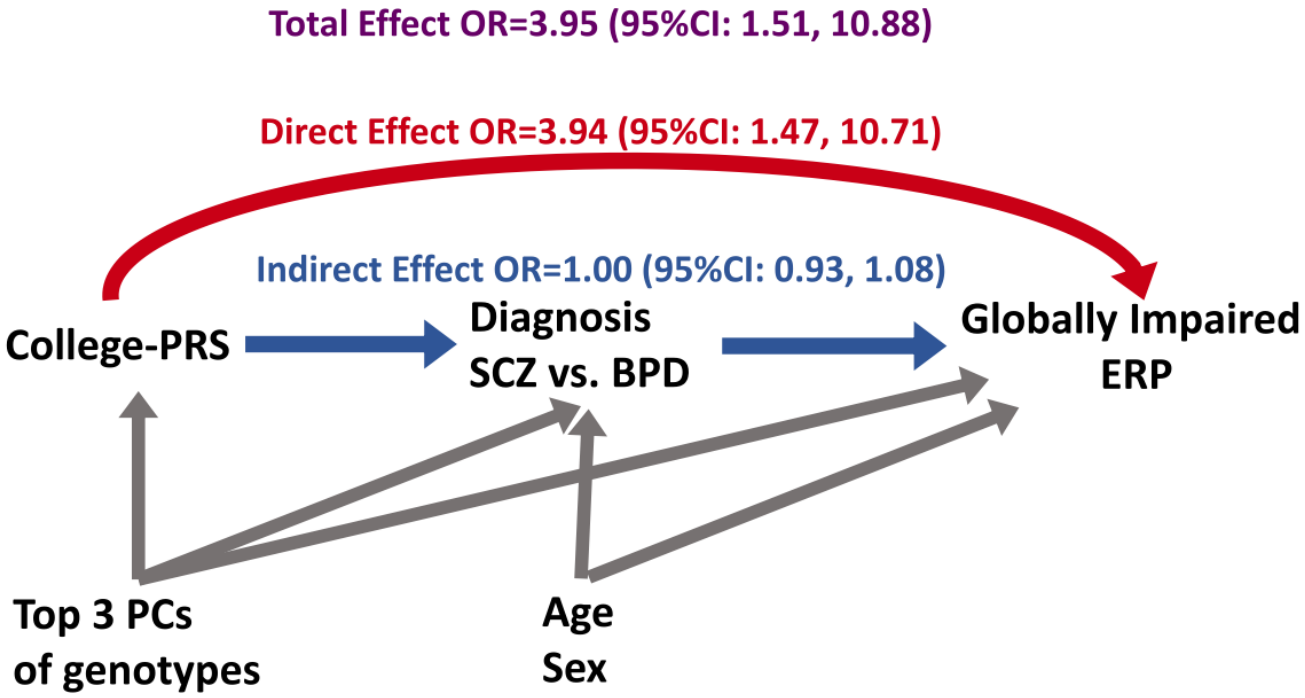
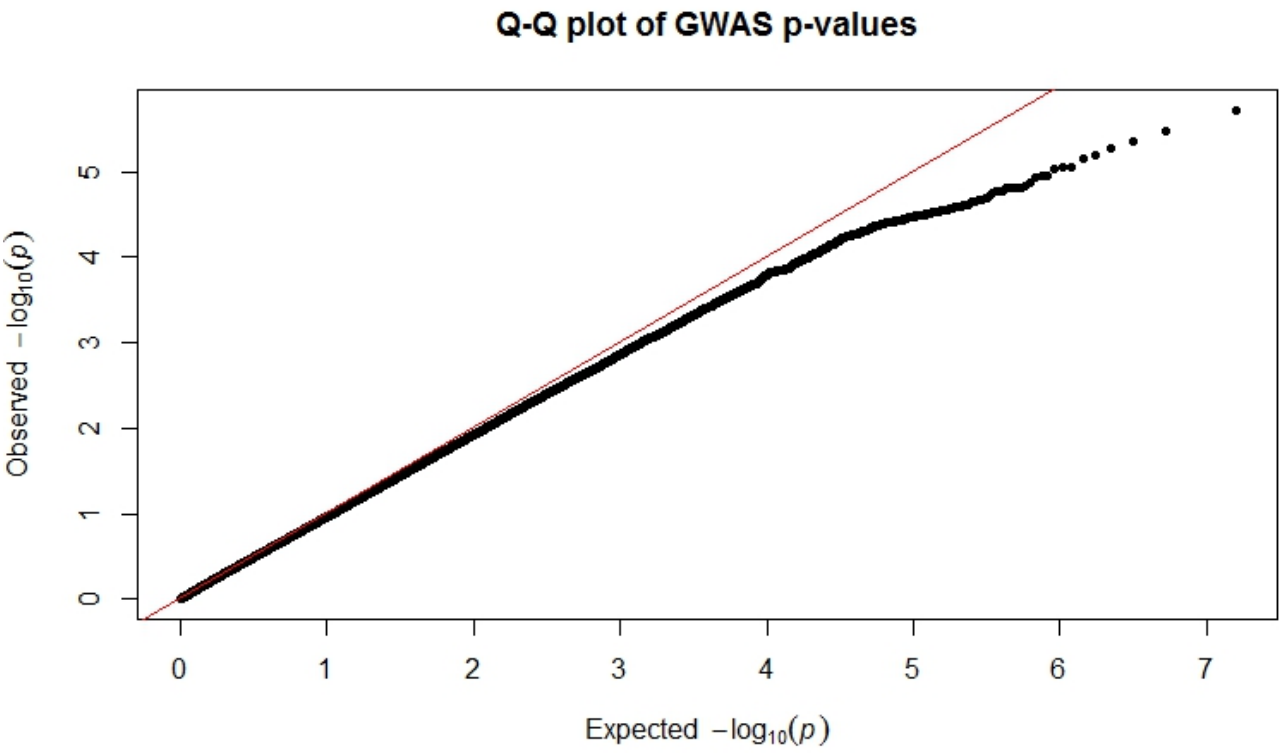
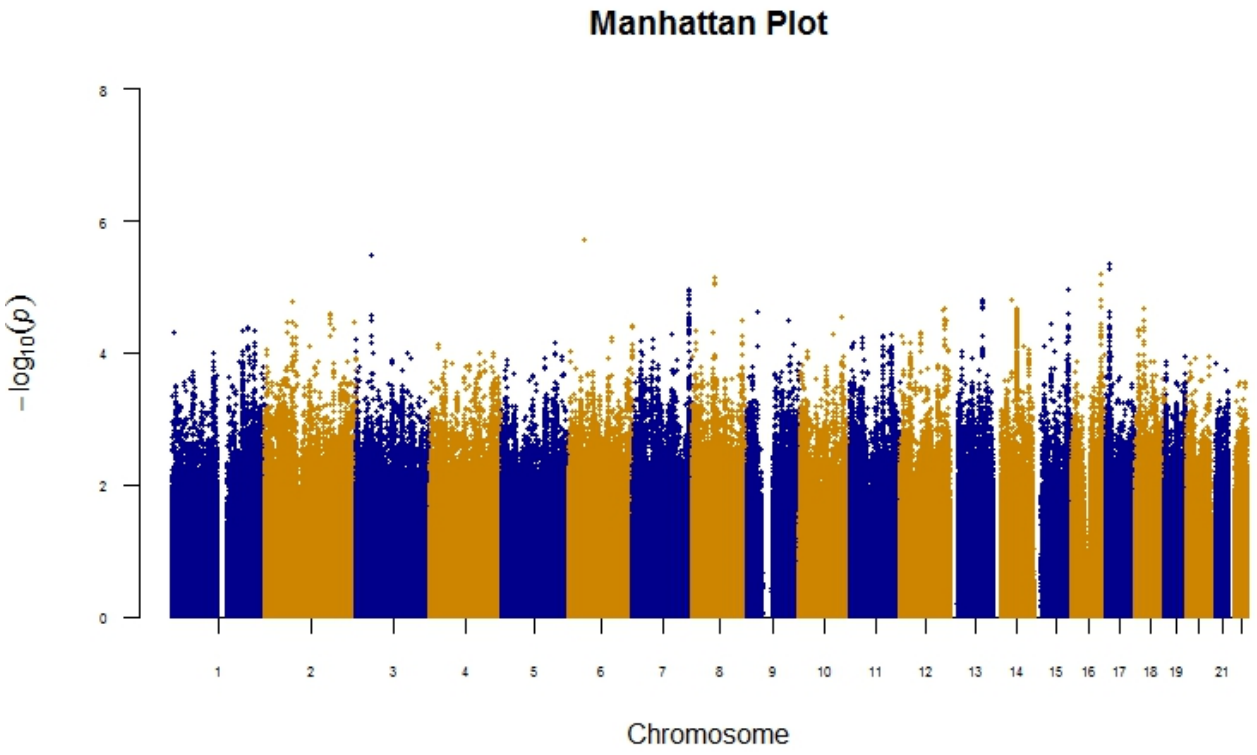


Figure S1.3: Quantile-quantile (Q-Q) plot of observed versus expected P values of the GWAS results for globally impaired ERP profile.



The straight line in the Q-Q plot indicates the distribution of SNPs under the null hypothesis.

Figure S1.4: Manhattan plot of the GWAS for globally impaired ERP profile.



P values (–log10) are plotted against their respective positions on each chromosome.

REFERENCES:

Cross-Disorder Group of the Psychiatric Genomics, C. (2013). Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet, 381*(9875), 1371-1379. doi:10.1016/S0140-6736(12)62129-1

Donchin, E., & Coles, M. G. H. (1988). Is the P300 Component a Manifestation of Context Updating. *Behavioral and Brain Sciences, 11*(3), 357-374.

Dudbridge, F. (2013). Power and predictive accuracy of polygenic risk scores. *PLoS Genet, 9*(3), e1003348. doi:10.1371/journal.pgen.1003348

Gratton, G., Coles, M. G., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalogr Clin Neurophysiol, 55*(4), 468-484.

Hall, M. H., Schulze, K., Rijsdijk, F., Kalidindi, S., McDonald, C., Bramon, E., . . . Sham, P. (2009). Are auditory P300 and duration MMN heritable and putative endophenotypes of psychotic bipolar disorder? A Maudsley Bipolar Twin and Family Study. *Psychol Med, 39*(8), 1277-1287. doi:S0033291709005261 [pii]

10.1017/S0033291709005261

Hall, M. H., Schulze, K., Rijsdijk, F., Picchioni, M., Ettinger, U., Bramon, E., . . . Sham, P. (2006). Heritability and Reliability of P300, P50 and Duration Mismatch Negativity. *Behav Genet, 36*(6), 845-857. doi:10.1007/s10519-006-9091-6

Hall, M. H., Taylor, G., Salisbury, D. F., & Levy, D. L. (2011). Sensory gating event-related potentials and oscillations in schizophrenia patients and their unaffected relatives. *Schizophr Bull, 37*(6), 1187-1199. doi:10.1093/schbul/sbq027

Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., . . . Benjamin, D. J. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature, 533*(7604), 539-542. doi:10.1038/nature17671

Polich, J., & Kok, A. (1995). Cognitive and biological determinants of P300: an integrative review. *Biol Psychol, 41*(2), 103-146.

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet, 38*(8), 904-909. doi:10.1038/ng1847

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., . . . Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet, 81*(3), 559-575. doi:S0002-9297(07)61352-4 [pii]

10.1086/519795

Rodriguez-Santiago, B., Brunet, A., Sobrino, B., Serra-Juhe, C., Flores, R., Armengol, L., . . . Perez-Jurado, L. A. (2010). Association of common copy number variants at the glutathione S-transferase genes and rare novel genomic changes with schizophrenia. *Mol Psychiatry, 15*(10), 1023-1033. doi:10.1038/mp.2009.53

Salisbury, D. F., Collins, K. C., & McCarley, R. W. (2010). Reductions in the N1 and P2 auditory event-related potentials in first-hospitalized and chronic schizophrenia. *Schizophr Bull, 36*(5), 991-1000. doi:10.1093/schbul/sbp003

Salisbury, D. F., Shenton, M. E., & McCarley, R. W. (1999). P300 topography differs in schizophrenia and manic psychosis. *Biol Psychiatry, 45*(1), 98-106.

Vanderweele, T. J. (2015). *Explanation in causal inference : methods for mediation and interaction.* . New York: Oxford University Press.

Chapter 2

Genetic overlap between vascular pathologies and Alzheimer's dementia and potential causal mechanisms

Yen-Feng Lin, Alkes Price, Jordan W. Smoller, Rebecca Betensky, and Deborah Blacker

ABSTRACT

OBJECTIVE:

Both cerebro- and cardio-vascular disease (collectively "CVD") may be genetically correlated with Alzheimer's dementia (AD) and late-life cognitive function. In this study, we examined the genetic overlap between vascular pathologies and AD dementia, and explored the extent to which vascular pathologies mediate the impact of AD dementia-related genetic variants on late-life cognition.

METHODS:

For 2,907 stroke- and dementia-free older individuals with genotype data available in the Age Gene/Environment Susceptibility (AGES)-Reykjavik Study, we generated genome-wide polygenic risk scores for AD dementia (GW-ADPRS), as well as *APOE*-linkage region polygenic risk scores (19q13-ADPRS), and non-*APOE* polygenic risk scores (non-19q13-ADPRS). We then examined the association of each PRS with markers of vascular pathology and cognitive function. Where associations were observed, we performed mediation analyses to identify whether association between the specific PRS and cognitive function was mediated by a vascular pathology.

RESULTS:

Memory score was significantly associated with GW-ADPRS (R^2=0.0029, P=0.001) and non-19q13-ADPRS (R^2=0.0020, P=0.008), and was nominally associated with 19q13-PRS (R^2=0.0015, P=0.02). GW-ADPRS was nominally associated with global cognition score (R^2=0.0015, P=0.02),

multiple lobar cerebral microbleeds (CMB) (R^2=0.0030, P=0.006), white matter lesion loads (WML) (R^2=0.0018, P=0.03) and coronary artery calcification (CAC) (R^2=0.0014, P=0.03); these genetic associations were almost totally explained by SNPs in the 19q13 region. The total effect of 19q13-ADPRS on memory was mediated by CMB (7.4%) and WML (5.6%), and its effect on global cognition was mediated by CMB (11.3%), WML (4.8%), and CAC (5.0%). The effect of non-19q13-ADPRS on both memory and global cognition was mediated by CMB (2.1% and 7.5%, respectively).

CONCLUSION:

Our findings support the hypothesis of a genetic overlap, mostly due to *APOE*, between vascular pathologies and AD dementia. The polygenic genetic effect on late-life cognition is partially but significantly mediated by CMB, WML, and CAC.

INTRODUCTION

Alzheimer's disease (AD) dementia is a major cause of morbidity and mortality in older adults (Daviglus et al., 2010). The etiology of AD dementia is complex and multifactorial. AD dementia refers to the clinical diagnosis of dementia considered likely to be due to underlying AD pathology. The major pathological hallmarks of AD are the accumulation of amyloid plaques and neurofibrillary tangles, which may lead to neurodegeneration and neuronal cell death (Mucke, 2009). However, it is well-established that a large fraction of those with a diagnosis of AD dementia also have cerebrovascular pathology, including white matter disease, lacunar and other small strokes, and microhemorrhages (Schneider, Arvanitakis, Bang, & Bennett, 2007; Toledo et al., 2013). Large clinically-recognized strokes can also be present, but they are more likely to have ruled out a diagnosis of "AD dementia," at least if they occur early in the course of disease. Systematically collected cohort-based autopsy data have shown that vascular pathology often coexists with AD pathology, adds to the likelihood of cognitive impairment, and lowers the threshold of AD pathology for the development of clinically diagnosed AD dementia (Schneider, Arvanitakis, Leurgans, & Bennett, 2009; Schneider & Bennett, 2010; Schneider, Wilson, Bienias, Evans, & Bennett, 2004).

A variety of cerebral small vessel diseases have been associated with AD dementia. Cerebral microbleeds (CMB) have been reported to be associated with the clinical manifestation and biochemical hallmarks of AD dementia (Goos et al., 2010; Goos et al., 2009; Vernooij et al., 2008). The presence of CMB in basal ganglia, thalamus, brainstem, and cerebellum is generally ascribed to hypertensive vasculopathy, while a lobar distribution of CMB is associated with

cerebral amyloid angiopathy (CAA) (Greenberg et al., 2009), which has been considered as a major contributor of the pathogenesis of AD dementia (Cordonnier & van der Flier, 2011). White matter lesions (WML), an imaging marker of cerebral small vessel disease, may also play a role in the development of AD dementia (Bilello et al., 2015; Inaba et al., 2011; Prins et al., 2004). A meta-analysis found that WML predicted an increased risk of AD dementia, all-type dementia, and stroke (Debette & Markus, 2010). Retinal imaging, a recently emerging tool to study cerebral small vessels in vivo, has been related to white matter lesions, brain atrophy, and an increased risk of dementia (de Jong et al., 2011; Ikram et al., 2006; Ikram et al., 2013).

Research efforts have also been devoted to the association between large vessel disease and AD dementia. Possible mechanisms linking large vessel atherosclerosis to AD dementia include shared etiology and brain hypoperfusion (de la Torre, 2010). Several longitudinal studies suggest that carotid intima-media thickness (CIMT), a marker of atherosclerosis, is associated with a later incidence of AD and dementia (Newman et al., 2005; van Oijen et al., 2007; Wendell et al., 2012). Additionally, a higher baseline CIMT and a more rapid progression of carotid atherosclerosis may accelerate cognitive decline in patients with AD dementia (Silvestrini et al., 2009). Atherosclerotic coronary artery calcification (CAC) is another useful marker of large vessel disease. Although there have been few reports on the relation between CAC and AD dementia, current evidence suggests that larger volume of CAC is associated with brain atrophy, loss of white matter microstructural integrity, and worse cognitive function in non-demented individuals (Bos et al., 2012; Reis et al., 2013).

Not surprisingly, epidemiological studies using clinical AD dementia as the outcome show that multiple vascular risk factors increase risk for AD (Weuve, 2017). As many risk factors for CVD are modifiable, focusing on the elucidation of the relationships and underlying mechanisms between vascular pathology and AD dementia might provide insight into the prevention or delay of AD-related cognitive decline in older individuals (de la Torre, 2010; Deschaintre, Richard, Leys, & Pasquier, 2009; Middleton & Yaffe, 2009).

To further investigate the etiological significance of CVD risk factors in AD dementia, genetic studies may also provide clues to understand the biological link between CVD and AD. Apolipoprotein E (*APOE*), the major susceptibility gene for AD (Michaelson, 2014), has been reported to be a risk factor for hyperlipidemia, cerebral lobar microbleeds, white matter lesions, ischemic stroke, and coronary heart disease (Bennet et al., 2007; Khan et al., 2013; Schilling et al., 2013). In addition to *APOE*, genome-wide association studies (GWAS) of AD dementia have identified single nucleotide polymorphisms (SNPs) with known or hypothesized relationships to lipid metabolism, such as *CLU* (clusterin), *ABCA7* (ABC transporter member 7), and *SORL1* (sortilin-related receptor) (Schellenberg & Montine, 2012). Recent studies using large-scale GWAS data suggest that AD dementia may be genetically correlated with levels of biomarkers for CVD risk (plasma lipids and C-reactive protein) (Desikan et al., 2015) and small vessel stroke (Traylor et al., 2016). A gene-based pathway approach to GWAS data has also identified shared genetic pathways between CVD and AD dementia (G. Liu et al., 2014).

A recent study found that the effect of *APOE*-e4, the AD-associated allele of *APOE*, on late-life cognition was partially mediated by cerebrovascular pathologies (Sajeev, 2015). In the present study, we expand to additional vascular pathologies beyond the brain and full genome data to more fully understand the relationship of AD genes and vascular pathology in the development of cognitive impairment. We generated genome-wide polygenic risk scores for AD dementia (GW-ADPRS) to examine the polygenic overlap between AD dementia and each of the following vascular pathologies: lobar CMB, WML, retinal venular diameter (RVD), CIMT, and CAC. We also generated two partitioned ADPRS, estimating genetic risk for AD dementia contributed separately by the 19q13 region that includes *APOE* and SNPs in linkage disequilibrium (LD) with *APOE*, and all other SNPs outside of the *APOE*-linkage region. We tested each ADPRS separately for association with cognition scores and with each vascular pathology. For vascular markers observed to be genetically correlated with AD dementia, we further performed formal mediation analyses to explore the causal relationship among genome-wide and 19q13 region ADPRSs, vascular pathology, and cognitive function.

METHODS

Study Sample Overview

The analyses were performed in data from the Age, Gene/Environment Susceptibility–Reykjavik Study (AGES–Reykjavik), a population-based epidemiologic study of genetic, behavioral and environmental risk factors for late-life disease and disability in the vascular, neurocognitive, musculoskeletal, and metabolic systems. The ascertainment and evaluation of the AGES sample has been described elsewhere (Harris et al., 2007). In brief, the AGES-Reykjavik sample was drawn from the Reykjavik Study, a population-based cohort study following up 30,795 randomly sampled Reykjavik residents born between 1905 and 1935. Between 2002 and 2006, 5,764 surviving participants of the Reykjavik Study were enrolled and evaluated with a questionnaire and various clinical, biochemical, and imaging exams. From the full 5,764 participants in AGES-Reykjavik, a random sample of 3,664 were selected for GWAS, of whom were excluded based on sample failure, genotype mismatch with reference panel, and sex mismatch, resulting in clean genotype data on 3,219 individuals.

For our phenotypic analyses, from the full AGES-Reykjavik sample of 5,764 participants, we excluded those with a history of stroke or vascular dementia, leaving 5,161. Of these participants, the 2,907 with clean genotype data available constituted the sample for our genetic analyses (see Figure 2.1).

AGES-Reykjavik was approved by the Icelandic National Bioethics Committee (VSN 00-063) and the Institutional Review Board of the U.S. National Institute on Aging. A document of informed consent was signed by all participants.

Genotyping and Quality Control

Genotyping was performed using the Illumina HumanCNV370-Duo (Illumina Inc.; San Diego, CA, USA) at the Laboratory of Neurogenetics, Intramural Research Program, at the U.S. National Institute of Aging. Standard protocols for working with Illumina data were followed, with a clustering score greater than 0.4. Prior to genotype imputation, we removed SNPs with call rate < 97%, Hardy-Weinberg Equilibrium < 1E-06, mishap < 1E-09, and mismatched positions between Illumina, dbSNP and/or HapMap. The quality control steps resulted in a total of 325,094 autosomal SNPs.

Genotype imputation was performed using the MaCH/ minimac program (Howie, Fuchsberger, Stephens, Marchini, & Abecasis, 2012), with the 1,000 Genomes Project dataset (Build: Hg19, 1000G v3 phase I) as the reference panel. We further filtered imputed SNPs based on imputation quality > 0.6, and the final imputed dataset consisted of 8,881,585 autosomal SNPs.

Markers of Vascular Pathologies

Lobar Cerebral Microbleeds

CMB were defined as focal hypointense lesions visible on the T2*-weighted GRE images, and lobar CMB were those located in the temporal, frontal, parietal, or occipital lobes. MRI

acquisition and reading protocol were described in detail previously (Qiu et al., 2008; Sveinbjornsdottir et al., 2008). CMB were assessed and recorded by neurologists and standardized raters according to previously reported criteria (Qiu et al., 2008; Sveinbjornsdottir et al., 2008). Multiple lobar CMB, an indicator of CAA (Greenberg et al., 2009), have been shown to be associated with decline in cognitive function in community older adults (Meier et al., 2014) or patients with stroke (Gregoire et al., 2012), and faster progression of dementia (Nagasawa, Kiyozaka, & Ikeda, 2014). In our analyses, we dichotomized lobar CMB as multiple (count > = 2) versus non-multiple (0 or 1).

Brain White Matter Lesion Load

WML were defined as visible hyperintense lesions on T2-weighted FSE/PD and FLAIR images. The whole brain total white matter lesion load was calculated as the summation of white matter hyperintensity volumes in the frontal, parietal, occipital, and temporal regions. The sample was grouped into quartiles of total white matter lesion load. In our analyses, individuals in the upper quartile of WML load were compared to those in the lower three quartiles. In a previous study using data from the AGES Reykjavik Study, participants in the upper quartile of WML load have been found to perform worse on cognitive tests (Saczynski et al., 2008).

Retinal Venular Diameter

Participants received digital retinal imaging of the fundus of each eye with a 6.3-megapixel Canon CR6 nonmydriatic digital camera (Klein et al., 2004; Qiu et al., 2008). All the retinal images were independently assessed by certified graders. The central retinal venular equivalent

(Knudtson et al., 2003) was calculated based on the measurement of the six largest venules within ½ to 1 disc diameter from the optic disc margin. The central retinal venular equivalent represents the average retinal venular diameter of both eyes. If one eye is not gradable, the value of the non-missing eye is used. This variable was treated as continuous in our analyses.

Carotid Intima–Media Thickness

Standard B-mode ultrasound images of the CIMT were acquired for the predefined segment of each of the right and left common carotid arteries at defined interrogation angles using the Meijers arc. Standard images were recorded from four angles at each site. The mean CIMT of the near and far walls were determined from a single image at each interrogation angle for both the right and left common carotid arteries. The details of the intima-media thickness analysis protocol were described elsewhere (Jonsson et al., 2009). We calculated the mean of all CIMT values and used the log-transformed mean of CIMT as a parameter in our analyses.

Coronary Artery Calcification

Images for calcium scoring were acquired using a Siemens Somatom Sensation 4 multi-detector Computed Tomography (CT) scanner (Siemens Medical Solutions, Erlangen, Germany) with prospective ECG triggering. Details of Coronary Computed Tomography Angiography acquisition and reading protocol were described elsewhere (Gudmundsson et al., 2012). Calcium in the coronary arteries was quantified using the Agatston scoring method (Agatston et al., 1990) by trained and certified image analysts. The total CAC score was calculated as the sum of the

scores for all four coronary arteries. We used the log-transformed total CAC score (log[CAC score+1] because there were zero values for the CAC score) as a variable in our analyses.

Cognitive function

Participants received a comprehensive cognitive assessment battery including tests of memory, executive function, and processing speed (Saczynski et al., 2008; Saczynski et al., 2009). Composite scores on the three cognitive domains for each subject were constructed by taking the mean of the Z-scores on individual tests within that domain. The composite memory score was calculated from scores on the California Verbal Learning Test (immediate and delayed recall); the composite executive function score was obtained from the Stroop Test (part 3), the CANTAB spatial working memory test, and the Digits Backward test; and the composite processing speed score was derived from scores on the Stroop Test (parts 1 and 2), Digit Symbol Substitution Test and Figure Comparison Test. The inter-rater reliability was high for all tests (Spearman correlation coefficients range from 0.96 to 0.99). The composite global cognition score was calculated as the mean of the three domain-specific composite scores. The two main outcomes used in our analyses were the standardized Z-score of the composite memory score and the standardized Z-score of the composite global cognition score.

Other Covariates

Other covariates used in the analyses included age (years, continuous), sex (binary), education (primary, secondary, college, university), smoking status (never, ever, or current), midlife physical activity (ideal, intermediate, or poor), diet quality (ideal, intermediate or poor),

prevalent diabetes (yes or no), hypertension (no, prehypertension, or hypertension), high LDL level (>=130 mg/dL), and obesity (BMI>=30). Levels of physical activity and diet quality were determined from brief questionnaires (Sturlaugsdottir et al., 2015) based on the American Heart Association guidelines on cardiovascular health (Lloyd-Jones et al., 2010). Diabetes was defined as self-reported doctor's diagnosis of diabetes, use of antidiabetic medication, or with a fasting glucose level >7mmol/L. Prehypertension was defined as a systolic blood pressure (SBP) from 120 to 139 millimeters of mercury (mmHg) or a diastolic blood pressure (DBP) from 80 to 89 mmHg, while hypertension was defined as SBP>=140 mmHg, DBP >=90 mmHg, doctor's diagnosis, or use of antihypertensive medication.

Polygenic Risk Scores (PRS) for AD Dementia and Associations with Cognitive or Vascular Phenotypes

Genome-wide polygenic risk score for AD dementia (GW-ADPRS)

We used the summary statistics from the Alzheimer's Disease Genetics Consortium (ADGC) GWAS (8,309 AD cases and 7,366 controls of European ancestry) (Naj et al., 2011) as the discovery dataset to derive genome-wide polygenic risk scores for AD dementia (GW-ADPRS) in our study sample. To include only independent association signals from the ADGC GWAS, we applied an LD clumping procedure to the discovery datasets: we retained the SNP with smallest P-value in each 250 kb window and removed all those in LD (r^2> 0.2) with this SNP. We used four different association P-value thresholds (P_Ts), 0.00001, 0.0001, 0.001, and 0.01, to select index SNPs from the clumped independent SNPs for generating the PRSs. For each individual, and each P-value threshold (P_T), we calculated GW-PRS by summing the risk allele counts of the

index SNPs, weighted by the log of their association odds ratios estimated from the ADGC GWAS results.

19q13-PRS and non-19q13-PRS for AD dementia

Because *APOE* is the strongest risk gene for AD dementia, we further partitioned the GW-ADPRS into an *APOE region* score and a non-*APOE* region score to separately assess the polygenic effects of SNPs in the *APOE*-linkage region 19q13 (ch19: 4500000-4580000) and all other SNPs. We followed the same steps as for the calculation of the GW-ADPRS to generate a 19q13-ADPRS (the summation of log-odds ratio weighted risk allele counts of the index SNPs in the 19q13 region) and a non-19q13-ADPRS (the summation of log odds ratio weighted risk allele counts of the index SNPs across whole genome except the 19q13 region) for each individual. The 19q13-ADPRS aggregates the effects of AD dementia-associated SNPs in the *APOE*-linkage region. Despite the clumping procedures, some of these SNPs may merely be tagging SNPs in LD with *APOE*, and others (*TOMM40*, *APOC1*, and *CD33*) may contribute independently to the risk of AD dementia. Therefore, we considered the 19q-13-ADPRS as an *APOE*-tagging-ADPRS rather than simply an *APOE* genetic risk score. However, because the *APOE* effect is very large, we expected that the effects of an *APOE* genetic risk score would be similar.

Data Analysis

Phenotypic associations of vascular markers with cognitive function

We used univariate and multivariate linear regressions to assess the associations of each vascular marker with the Z-score of the composite memory score and the Z-score of the composite global cognition score. Multivariate models adjusted for age, sex, education, diabetes, hypertension, high LDL level, obesity, physical activity, diet quality, and smoking status.

Association of ADPRS with vascular and cognitive phenotypes

We then performed PRS association analyses to examine if any of the P_Ts generates an ADPRS significantly associated with the following phenotypes: The Z-score of the composite memory score, the Z-score of the composite global cognition score, and each of the vascular pathologies (CMB, WML, RVD, CIMT, and CAC). Associations were tested using linear (for continuous phenotypes) or logistic (for binary phenotypes) regression models with baseline age and sex as covariates. We tested the association of each phenotype with each of GW-ADPRS, 19q13-ADPRS, and non-19q13-ADPRS, adjusting for age and sex. The Wald test P-value for each association test was reported, and squared semi-partial correlations (R^2) were calculated to estimate the proportion of variance explained by the PRSs.

Causal Mediation Analyses

We then explored the causal relationship between vascular pathologies, late-life cognition, and the PRS for AD dementia that might explain any observed genetic correlation among these factors. We performed causal mediation analyses to examine how much of the effect of an ADPRS on cognition score was mediated by a vascular pathology observed to be genetically

correlated with AD dementia. We used a method based on the counterfactual framework for causal inference (Pearl, 2001; Robins & Greenland, 1992), which is an extension of traditional regression-based mediation approaches (Baron & Kenny, 1986), allowing binary mediators and outcomes, as well as exposure-mediator interactions (T. J. VanderWeele, 2016).

For each of the ADPRSs (GW-ADPRS, 19q13-ADPRS, and non-19q18_ADPRS) as the exposure, we estimated the direct and indirect (mediated) effects of each vascular pathology as the mediator, and Z-score of the composite memory or global cognition score as the outcome. In order to gain more statistical power, the ADPRS exposures used in the mediation analyses were those with a P-value threshold that showed the highest association with each cognitive outcome. We adjusted for potential mediator-outcome confounders, including age, sex, smoking status, midlife physical activity, diet quality, and other genetic risk scores if necessary. A counterfactual outcome variable denotes the outcome that would have been observed had an exposure been set to a particular value. In order to compare high and low values of each ADPRS in our estimates of the direct and the indirect effects we chose to compare the 75th percentile and the 25th percentile of each.

All the mediation analyses were performed by using the PARAMED module in STATA (Emsley & Liu, 2013; Valeri & Vanderweele, 2013). We used bootstrap procedures with 200 replications to compute a 95% bias-corrected bootstrap confidence interval (95% BCCI) for the direct and indirect effects.

Finally, we conducted sensitivity analyses. First, we followed the procedures developed by Imai et al. (K. Imai, Keele, & Tingley, 2010; K. K. Imai, L.; Yamamoto, T., 2010) to evaluate the robustness of the above mediation analyses to unmeasured confounding (see Supplementary Methods). Then as a sensitivity analysis of the 75^{th} vs. 25^{th} percentile comparison, we also repeated all mediation analyses comparing the effects of the 90th percentile and the 10th percentile of each ADPRS.

Correction for Multiple Testing

In the PRS association analyses, we considered a PRS-wise significant threshold for the correction of multiple comparisons ($P< 0.008$, after Bonferroni correction for the 6 association tests between 2 cognitive outcomes and 3 ADPRS for each genomic region; and $P<0.003$, after Bonferroni correction for the 15 association tests between 5 vascular pathologies and 3 ADPRS for each genomic region). Bonferroni correction is conservative here, because the vascular and cognitive phenotypes are somewhat correlated.

In the mediation analyses, vascular pathologies with a p-value lower than 0.05 for PRS associations with AD dementia were tested as potential mediators. While multiple testing corrections should be considered when interpreting results of the PRS association analyses, we used a nominal significance of 0.05 for our purposes, as we were more concerned that we did not miss traits that are genetically correlated with AD dementia but did not achieve statistical significance due to a small sample size.

RESULTS

Sample Characteristics

Table 2.1 presents descriptive statistics of baseline characteristics for the AGES sample used here. The mean age of all stroke- and dementia-free subjects (n=5,161) was 76.7 (5.8) years. Approximately 80% of subjects had hypertension, but vascular pathologies were relatively rare: for example, only 2% had multiple lobar CMB. Subjects with genotype data available (2,907) were very similar to the full sample, as would be expected, but those with had somewhat lower coronary calcification score (P=0.01).

Phenotypic Associations

Table 2.2 presents phenotypic associations between each vascular pathology and cognitive outcomes. All unadjusted associations were significant. After adjusting for potential confounders, CMB, CAC, and WML were significantly associated with memory score, whereas the former two were significantly associated with global cognition score. If all potential confounders remain constant, those with multiple lobar CMB averagely had 0.19 SD lower memory score and 0.26 SD lower global cognition score, than those with 0 or 1 lobar CMB. Individuals in the highest quartile of WML load had 0.07 SD lower memory score than others. For each one-unit increase in the log-transformed CAC, the mean memory and global cognition scores declined by 0.015 SD and 0.023 SD, respectively.

Associations of GW-ADPRS with Cognitive or Vascular Phenotypes

We performed PRS association analyses to examine the association between GW-ADPRS and each vascular pathology, and the association between GW-ADPRS and each cognitive outcome. The results of PRS association analyses are presented in Figure 2.2 and Table 2.3.

Memory score was significantly associated with the GW-ADPRS with a P-value threshold of 0.0001 (GW-ADPRS$_{PT = 0.0001}$; P = 0.006, R^2 = 0.22%) and the GW-ADPRS with a P-value threshold of 0.01 (GW-ADPRS$_{PT = 0.01}$; P = 0.001, R^2 = 0.29%), after the Bonferroni correction for multiple comparisons. We also found nominal associations of global cognition score with GW-ADPRS$_{PT = 0.0001}$ and GW-ADPRS$_{PT = 0.01}$. In terms of the association between the GW-ADPRSs and vascular pathologies, we found that the GW-ADPRS with at all three P_T were nominally associated with multiple lobar CMB. There were also nominal associations between GW-ADPRS$_{PT = 0.01}$ and WML and between GW-ADPRS$_{PT = 0.0001}$ and CAC.

Associations of 19q13-PRS and non-19q13-PRS with Cognitive or Vascular Phenotypes

Using the two partitions of the GW-ADPRS, the 19q13-ADPRS and the non-19q13-ADPRS at each P threshold, we tested for association with each cognitive and vascular phenotype (Table 2.3). Our data showed that the 19q13-ADPRS at all three P thresholds was significantly associated with lobar CMB. We also found nominal associations of 19q13-PRS with WML, CAC, and both cognitive outcomes. For non-19q13-PRS, the only association was that between non-19q13-PRS$_{PT = 0.01}$ and memory score.

Mediation Analyses

Our PRS association analyses suggested possible SNP-based genetic overlap of AD dementia with CMB, WML, and CAC. To examine whether the effect of ADPRS on late-life cognition was mediated through these vascular pathologies, we performed regression-based causal mediation analyses to decompose the total effect of each PRS on each of the cognitive outcomes (i.e., Z-score of the composite memory score and Z-score of the composite global cognition score) into direct and indirect (mediated) effects, adjusting for potential confounders. In these mediation analyses, the PRS that most associated with each cognitive outcome was selected as the exposure (GW-ADPRS$_{PT = 0.01}$, 19q13-ADPRS$_{PT = 0.001}$, and non-19q13-ADPRS$_{PT = 0.01}$ for memory; GW-ADPRS$_{PT = 0.0001}$, 19q13-ADPRS$_{PT = 0.001}$, non-19q13-ADPRS$_{PT = 0.01}$ for global cognition). Vascular pathologies with a p-value lower than 0.05 for PRS associations with AD dementia were tested as potential mediators.

Relationship between ADPRS, Vascular Pathologies, and Cognitive Decline

We estimated the direct and indirect effects on memory and global cognition scores, comparing the 75th percentile versus the 25th percentile of the each ADPRS, adjusting for potential confounders. Results are shown in Table 2.4. The proportion mediated (PM) was obtained by dividing the estimated indirect effect by the estimated total effect, as an index of the degree of mediation. The total effect of GW-ADPRS on memory score was significantly mediated by multiple lobar CMB (PM= 3.2% [95%CI= 0.7%, 7.8%]) and WML load (PM=2.7% [95%CI= 0.05%, 8.9%]). Of the five vascular pathologies tested, CMB (PM=10.8% [95%CI= 4.0%, 24.1%]), WML (PM=4.0% [95%CI= 0.7%, 15.2%]), and CAC (PM=4.4% [95%CI= 0.2%, 12.4%]) were identified as significant mediators of the effects of GW-ADPRS on global cognition score. The total effect of

19q13-ADPRS on memory score was significantly mediated by CMB (PM=7.4% [95%CI= 1.3%, 19.9%]) and WML (PM=5.6% [95%CI= 0.9%, 18.4%]), and its effect on global cognition was mediated by CMB (PM=11.3% [95%CI= 3.6%, 35.3%]), WML (PM=4.8% [95%CI= 0.5%, 17.3%]), and CAC (PM=5.0% [95CI= 0.1%, 13.2%]). The total effect of non-19q13-ADPRS on both memory and global cognition was mediated by CMB (PM= 2.1% [95%CI= 0.1%, 6.8%] and 7.5% [95%CI= 0.2%, 22.6%], respectively).

When an interaction between the PRS and the marker of vascular pathology was included in each mediation model, there was little or no change in the estimated direct and indirect effects, so we decided not to include the interaction in the mediation models, as suggested by Vanderweele (T.J. Vanderweele, 2015).

Sensitivity Analyses

Sensitivity analyses of unmeasured confounding suggest that in the presence of unmeasured confounders (or residual confounding) associated with better cognition and less severe vascular pathology or unmeasured confounders associated with poorer cognition and more severe vascular pathology, our estimated PMs would underestimate the true mediation effects of vascular pathologies. Under the seemingly less likely scenarios of unmeasured confounders associated with better cognition and more severe vascular pathology, or unmeasured confounders associated with poorer cognition and less severe vascular pathology, our estimated PMs would overestimate the true mediation (see Supplementary Results and Figure S2.1). The degree of unmeasured confounding was estimated by the size of the correlation

(denoted by rho) between residuals in the equation predicting the mediator and the equation predicting the outcome. For example, in the mediation analysis with GW-ADPRS as the exposure, CMB as the mediator, and memory score as the outcome, the PM would reduce from 3.2% to 0 if rho equals to 0.13 and would be greater than 3.2% if rho is less than 0. The rho which would reduce the PM to 0 for each mediation model ranges from 0.03 to 0.21 (see Supplementary Results and Table S2.2).

Sensitivity analyses for selection of exposure levels for comparison found that mediation analyses comparing the effects of the 90th percentile and the 10th percentile of each PRS yielded very similar PMs as those shown in Table 2.4 (see Supplementary Table S2.1).

DISCUSSION

In a community-based sample of 5,161 stroke- and dementia-free older individuals, we found that multiple lobar cerebral microbleeds, higher white matter lesion load, and greater coronary artery calcification—but not retinal venular diameter nor carotid intima-media --were associated with poorer late-life memory and global cognition. In the 2,907 genotyped individuals, we found that a higher genetic risk score for AD dementia was associated with these three vascular pathologies and two cognition outcomes. In mediation analyses, we found that the effects of *APOE* and SNPs near *APOE* on memory may be partially mediated by CMB and WML, and their effects on global cognition may be partially mediated by CMB, WML, and CAC. With the possible exception of CMB, there was little evidence of an effect of non-*APOE* AD dementia-associated alleles on either memory or global cognition.

Genetic Overlap between AD Dementia and Vascular Pathologies

In a relatively large sample of older adults, our phenotypic analyses replicated previously reported associations of cerebral small vessel disease (Poels et al., 2012; Staals et al., 2015) and atherosclerosis (Bos et al., 2012) with cognitive function. We then further examined if shared genetic factors contribute to these observed associations. Our results demonstrate genetic overlap between AD dementia, vascular pathologies, and cognitive function in older individuals free of prevalent dementia or stroke (Table 2.2 and Figure 2.2). We found an association between the GW-ADPRS and multiple lobar CMB. The 19q13-ADPRS was even more strongly associated (significant after Bonferroni correction) with CMB, suggesting AD dementia associated SNPs within and surrounding *APOE* were associated with having multiple lobar CMB.

Similar associations were also observed for GW-ADPRS and 19q13-ADPRS with WML load and CAC score, but no association was observed either for RVD or CIMT. Our findings of the strongest genetic overlap between AD dementia and CMB are consistent with a recent reported genetic correlation between AD and cerebral small vessel disease but not large vessel disease (Traylor et al., 2016). Lobar CMB may be caused by CAA (Smith et al., 2010; Yates et al., 2014), which is highly prevalent in post-mortem analyses of brains of persons with a clinical diagnosis of AD dementia during life (Jellinger, 2002). In addition, the *APOE*-e4 allele has been associated with the presence of CAA (Esiri et al., 2015; Pfeifer, White, Ross, Petrovitch, & Launer, 2002). These previous findings may support the possible genetic overlap between lobar CMB and AD dementia observed in our data. Since WML (de Leeuw et al., 2002) and CAC (W. Liu et al., 2015) may share some common risk factors with CMB, and both have been related to dementia (although the evidence is not as strong as that for CMB), a genetic overlap of AD dementia with WML and CAC makes sense.

Although both CMB and WML were observed to be genetically correlated with AD, our results showed no association between the genetic risk score for AD dementia and RVD, which has been reported as an indicator of cerebral small vessel pathology (Ikram et al., 2006). One possible explanation is that the central retinal venular equivalent is a more observer-dependent measure and may not accurately reflect the degree of retinal venular dilatation, but there is no indication of even an attenuated signal in the beta coefficients or the squared semi-partial correlations (i.e., R^2, the proportion of variance in RVD measure explained by ADPRS).

APOE Explains the Genetic Overlap of AD Dementia and Vascular Pathologies

We separately examined *APOE* and non-*APOE* contributions to the association of the genome-wide polygenic risk score for AD dementia with vascular pathologies and late-life cognition. Our results showed that the *APOE* gene explains most of the SNP-based genetic overlap of AD dementia with the vascular pathologies. *APOE* has been related to cerebrovascular dysfunction by affecting cerebral blood flow, blood brain barrier (BBB) integrity, and neuronal-vascular coupling (Tai et al., 2016). As mentioned above, the *APOE*-e4 allele is a risk factor for CAA (Esiri et al., 2015; Pfeifer et al., 2002). In terms of peripheral vascular disease, *APOE* has been shown to be an important factor in the development of hyperlipoproteinemia and atherosclerosis (Huang & Mahley, 2014; Tai et al., 2016). Our data also showed *no association* between the non-19q13-ADPRS and any of the examined vascular pathologies, despite previously reported associations of non-APOE AD risk genes with inflammation and abnormal lipid metabolism, which are both risk factors for vascular disease (Karch & Goate, 2015). Future research with larger samples are needed to test for association between vascular pathologies and AD dementia-associated alleles outside of the APOE-linkage region.

In terms of the association between the ADPRS and performance in cognitive tests, in our sample, the *APOE* gene and its nearby SNPs contributed most of the genetic overlap between AD dementia and global cognition, whereas both the *APOE*- and non-*APOE*-ADPRS were associated with memory. These results were consistent with previous findings on the relationship between AD-associated SNPs and different domains of cognitive function. A large-scale meta-analysis including 77 studies of the association between *APOE* and cognitive

function suggested that carriers of *APOE*-e4 performed worse on multiple domains of cognitive tests, including episodic memory, executive functioning, perceptual speed, and overall global cognitive ability (Wisdom, Callahan, & Hawkins, 2011). On the other hand, non-*APOE*-ADPRS calculated by using summary statistics from the International Genomics of Alzheimer's Project (IGAP) was found to be associated with memory but not executive function in non-demented subjects in the Alzheimer's Disease Neuroimaging Initiative (ADNI) (Mormino et al., 2016). The ADNI sample used by Mormino et al., with mean age of 75.3 years, is very similar to ours. It is possible that impaired memory was more likely to be detected than deficit in other cognitive domains for individuals at this age (at this stage of cognitive decline).

Potential Causal Relationships between PRS for AD Dementia, Vascular Pathologies, and Cognition

Having established a SNP-based genetic overlap between AD dementia, vascular pathologies, and late-life cognitive function, we then sought to identify the causal relationships between PRS for AD dementia, vascular pathologies, and cognition scores. When considering 19q13-PRS as the exposure, we found that lobar CMB, WML load, and CAC significantly mediated the polygenic effects on global cognition, while lobar CMB and WML load mediated the effects on memory. When considering the non-19q13-ADPRS as the exposure, there was no evidence of mediation, except perhaps for lobar CMB. Our findings indicate that AD dementia-associated SNPs affect late-life cognition partially through pathways involving vascular pathologies, providing insight into potential pathogenic mechanisms in clinical AD dementia. The results also may lend further support to interventions to reduce vascular pathologies may be of value in the

prevention of AD dementia. It is worth noted that we separately examined the mediation effects of lobar CMB, WML load, and CAC. Although measures of these vascular pathologies were correlated with each other, their correlations were relatively weak in our sample (Kendall's tau-b = 0.07 for CMB-WML, Point-Biserial correlation coefficient = 0.08 for CMB-CAC and 0.12 for WML-CAC). Thus, it is reasonable to believe that a certain proportion of AD dementia-associated SNP effects on cognitive function was mediated by vascular pathologies when considering all vascular mediators together.

The only previous study investigating the mediation role of cerebrovascular imaging markers between genetic variants and cognitive function, which used an overlapping sample from the same cohort as ours (i.e., the AGES-Reykjavik), reported that about 9% of the total effect of *APOE*4 carriership on global cognition was mediated by CMB and WML volume (Sajeev, 2015). Our analyses revealed similar but stronger mediation effect of vascular pathologies on the relationship between SNPs and cognitive function. The major strength of the present study is that we assessed the effects of PRS, aggregating multiple possible risk alleles for AD across the whole genome, within or beyond the *APOE*-linkage region. Moreover, we considered both cerebral small vessel and systemic large vessel imaging markers, which have been previously associated with dementia or poor cognition, as potential mediators.

Possible Causal Relationship between Vascular Pathologies and Poorer Cognition

In our analyses of phenotypic associations between vascular pathologies and cognitive function (Table 2.2), although we have considered a number of potential confounders, we cannot

completely rule out the possibilities of residual confounding. With the established causal mediation relationship between PRS, vascular pathologies, and cognitive performance (Figure 2.3), PRS for AD dementia appears to affect both vascular pathologies and cognition and may confound the causal effect of vascular pathologies on cognition. In addition to the covariates included in the phenotypic analyses shown in Table 2.2, we also adjusted for the 19q13- and non-19q13-PRS that most associated with the cognitive outcomes (i.e., 19q13-$PRS_{PT = 0.001}$, and non-19q13-$PRS_{PT = 0.01}$) to account for possible residual confounding by shared genetic factors. In the fully-adjusted analyses, the effects from CMB and WML to cognitive outcomes remained significant; however, the associations between CAC and cognition were greatly attenuated and the association with memory was rendered non-significant after also accounting for the PRS (Table 2.5).

Limitations

Several limitations in the present study should be noted. In the population-based sample, in which most subjects were cognitively normal or mildly impaired, mean scores of cognitive tests reflect both lifelong cognitive variability and recent pathological changes, and the former may overwhelm the latter. However, with our relatively large sample size, we were able to detect small signals and parse these signals into what appear to be meaningful mediation relationships. Nonetheless, the sample may have only been large enough to detect APOE-related signals (even if other causal SNPs are present). In any event, in the setting of small signals, another major limitation of the present study is the possible violation of the no-unmeasured-confounding assumption necessary for causal mediation analyses. However, our

sensitivity analyses suggest that given the expected direction of unmeasured confounding, our estimated indirect effects may underestimate the true mediated effects. In addition, the genetic risk scores, including only common genetic variants, cannot account for all the genetic effects on cognitive performance and AD dementia. Although our SNP-based genetic risk scores were strongly associated with vascular and cognitive phenotypes, and PRS for AD dementia has been reported to be capable to capture nearly all common genetic risk for AD (Escott-Price, Shoai, Pither, Williams, & Hardy, 2017), there are still causal genomic variants (e.g., rare variants) that are not well-tagged by GWAS SNPs. However, the genetic effects not captured by SNP-based risk scores can also be seen as a type of unmeasured mediator-outcome confounding. Therefore, the sensitivity analyses mentioned above may help minimize these concerns. Finally, although using a sample from the relatively isolated and genetically homogeneous Icelandic population (Helgason, Nicholson, Stefansson, & Donnelly, 2003) may enhance the power of a genetic epidemiological study (Harris et al., 2007), it limits the generalizability of the findings to other ethnic populations. Future research including a broader range of ethnic populations is needed.

CONCLUSION

This is the first study, to our knowledge, that combined polygenic profiling and causal mediation methods to identify the causal relationship between two genetically correlated phenotypes and their shared genetic factors. Our findings support the hypothesis of a genetic overlap, mostly due to *APOE*, between AD dementia and vascular pathologies, especially small vessel disease. Our results also showed that in older individuals, CMB, WML, and CAC may causally affect cognitive function and partially mediate the polygenic genetic effects of AD-related genes on cognition, underscoring the potential role of vascular factors in cognitive decline, and suggesting vascular pathologies as a target for future mechanistic research in this area.

REFERENCES:

Agatston, A. S., Janowitz, W. R., Hildner, F. J., Zusmer, N. R., Viamonte, M., Jr., & Detrano, R. (1990). Quantification of coronary artery calcium using ultrafast computed tomography. *J Am Coll Cardiol, 15*(4), 827-832.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Pers Soc Psychol, 51*(6), 1173-1182.

Bennet, A. M., Di Angelantonio, E., Ye, Z., Wensley, F., Dahlin, A., Ahlbom, A., . . . Danesh, J. (2007). Association of apolipoprotein E genotypes with lipid levels and coronary risk. *JAMA, 298*(11), 1300-1311. doi:10.1001/jama.298.11.1300

Bilello, M., Doshi, J., Nabavizadeh, S. A., Toledo, J. B., Erus, G., Xie, S. X., . . . Davatzikos, C. (2015). Correlating Cognitive Decline with White Matter Lesion and Brain Atrophy Magnetic Resonance Imaging Measurements in Alzheimer's Disease. *J Alzheimers Dis, 48*(4), 987-994. doi:10.3233/JAD-150400

Bos, D., Vernooij, M. W., Elias-Smale, S. E., Verhaaren, B. F., Vrooman, H. A., Hofman, A., . . . Ikram, M. A. (2012). Atherosclerotic calcification relates to cognitive function and to brain changes on magnetic resonance imaging. *Alzheimers Dement, 8*(5 Suppl), S104-111. doi:10.1016/j.jalz.2012.01.008

Cordonnier, C., & van der Flier, W. M. (2011). Brain microbleeds and Alzheimer's disease: innocent observation or key player? *Brain, 134*(Pt 2), 335-344. doi:10.1093/brain/awq321

Daviglus, M. L., Bell, C. C., Berrettini, W., Bowen, P. E., Connolly, E. S., Jr., Cox, N. J., . . . Trevisan, M. (2010). NIH state-of-the-science conference statement: Preventing Alzheimer's disease and cognitive decline. *NIH Consens State Sci Statements, 27*(4), 1-30.

de Jong, F. J., Schrijvers, E. M., Ikram, M. K., Koudstaal, P. J., de Jong, P. T., Hofman, A., . . . Breteler, M. M. (2011). Retinal vascular caliber and risk of dementia: the Rotterdam study. *Neurology, 76*(9), 816-821. doi:10.1212/WNL.0b013e31820e7baa

de la Torre, J. C. (2010). Vascular risk factor detection and control may prevent Alzheimer's disease. *Ageing Res Rev, 9*(3), 218-225. doi:10.1016/j.arr.2010.04.002

de Leeuw, F. E., de Groot, J. C., Oudkerk, M., Witteman, J. C., Hofman, A., van Gijn, J., & Breteler, M. M. (2002). Hypertension and cerebral white matter lesions in a prospective cohort study. *Brain, 125*(Pt 4), 765-772.

Debette, S., & Markus, H. S. (2010). The clinical importance of white matter hyperintensities on brain magnetic resonance imaging: systematic review and meta-analysis. *BMJ, 341*, c3666. doi:10.1136/bmj.c3666

Deschaintre, Y., Richard, F., Leys, D., & Pasquier, F. (2009). Treatment of vascular risk factors is associated with slower decline in Alzheimer disease. *Neurology, 73*(9), 674-680. doi:10.1212/WNL.0b013e3181b59bf3

Desikan, R. S., Schork, A. J., Wang, Y., Thompson, W. K., Dehghan, A., Ridker, P. M., . . . DemGene, I. (2015). Polygenic Overlap Between C-Reactive Protein, Plasma Lipids, and Alzheimer Disease. *Circulation, 131*(23), 2061-2069. doi:10.1161/CIRCULATIONAHA.115.015489

Emsley, R., & Liu, H. (2013). PARAMED: Stata module to perform causal mediation analysis using parametric regression models: Boston College Department of Economics. Retrieved from https://ideas.repec.org/c/boc/bocode/s457581.html

Escott-Price, V., Shoai, M., Pither, R., Williams, J., & Hardy, J. (2017). Polygenic score prediction captures nearly all common genetic risk for Alzheimer's disease. *Neurobiol Aging, 49*, 214 e217-214 e211. doi:10.1016/j.neurobiolaging.2016.07.018

Esiri, M., Chance, S., Joachim, C., Warden, D., Smallwood, A., Sloan, C., . . . Smith, A. D. (2015). Cerebral amyloid angiopathy, subcortical white matter disease and dementia: literature review and study in OPTIMA. *Brain Pathol, 25*(1), 51-62. doi:10.1111/bpa.12221

Goos, J. D., Henneman, W. J., Sluimer, J. D., Vrenken, H., Sluimer, I. C., Barkhof, F., . . . van der Flier, W. M. (2010). Incidence of cerebral microbleeds: a longitudinal study in a memory clinic population. *Neurology, 74*(24), 1954-1960. doi:10.1212/WNL.0b013e3181e396ea

Goos, J. D., Kester, M. I., Barkhof, F., Klein, M., Blankenstein, M. A., Scheltens, P., & van der Flier, W. M. (2009). Patients with Alzheimer disease with multiple microbleeds: relation with cerebrospinal fluid biomarkers and cognition. *Stroke, 40*(11), 3455-3460. doi:10.1161/STROKEAHA.109.558197

Greenberg, S. M., Vernooij, M. W., Cordonnier, C., Viswanathan, A., Al-Shahi Salman, R., Warach, S., . . . Microbleed Study, G. (2009). Cerebral microbleeds: a guide to detection and interpretation. *Lancet Neurol, 8*(2), 165-174. doi:10.1016/S1474-4422(09)70013-4

Gregoire, S. M., Smith, K., Jager, H. R., Benjamin, M., Kallis, C., Brown, M. M., . . . Werring, D. J. (2012). Cerebral microbleeds and long-term cognitive outcome: longitudinal cohort study of stroke clinic patients. *Cerebrovasc Dis, 33*(5), 430-435. doi:10.1159/000336237

Gudmundsson, E. F., Gudnason, V., Sigurdsson, S., Launer, L. J., Harris, T. B., & Aspelund, T. (2012). Coronary artery calcium distributions in older persons in the AGES-Reykjavik study. *Eur J Epidemiol, 27*(9), 673-687. doi:10.1007/s10654-012-9730-6

Harris, T. B., Launer, L. J., Eiriksdottir, G., Kjartansson, O., Jonsson, P. V., Sigurdsson, G., . . . Gudnason, V. (2007). Age, Gene/Environment Susceptibility-Reykjavik Study: multidisciplinary applied phenomics. *Am J Epidemiol, 165*(9), 1076-1087. doi:10.1093/aje/kwk115

Helgason, A., Nicholson, G., Stefansson, K., & Donnelly, P. (2003). A reassessment of genetic diversity in Icelanders: strong evidence from multiple loci for relative homogeneity caused by genetic drift. *Ann Hum Genet, 67*(Pt 4), 281-297.

Howie, B., Fuchsberger, C., Stephens, M., Marchini, J., & Abecasis, G. R. (2012). Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet, 44*(8), 955-959. doi:10.1038/ng.2354

Huang, Y., & Mahley, R. W. (2014). Apolipoprotein E: structure and function in lipid metabolism, neurobiology, and Alzheimer's diseases. *Neurobiol Dis, 72 Pt A*, 3-12. doi:10.1016/j.nbd.2014.08.025

Ikram, M. K., De Jong, F. J., Van Dijk, E. J., Prins, N. D., Hofman, A., Breteler, M. M., & De Jong, P. T. (2006). Retinal vessel diameters and cerebral small vessel disease: the Rotterdam Scan Study. *Brain, 129*(Pt 1), 182-188. doi:10.1093/brain/awh688

Ikram, M. K., de Jong, F. J., Vernooij, M. W., Hofman, A., Niessen, W. J., van der Lugt, A., . . . Ikram, M. A. (2013). Retinal vascular calibers associate differentially with cerebral gray matter and white matter atrophy. *Alzheimer Dis Assoc Disord, 27*(4), 351-355. doi:10.1097/WAD.0b013e31829344ed

Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. *Psychol Methods, 15*(4), 309-334. doi:10.1037/a0020761

Imai, K. K., L.; Yamamoto, T. (2010). Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science, 25*, 51-71. doi:10.1214/10-STS321

Inaba, M., White, L., Bell, C., Chen, R., Petrovitch, H., Launer, L., . . . Masaki, K. (2011). White matter lesions on brain magnetic resonance imaging scan and 5-year cognitive decline: the Honolulu-Asia aging study. *J Am Geriatr Soc, 59*(8), 1484-1489. doi:10.1111/j.1532-5415.2011.03490.x

Jellinger, K. A. (2002). Alzheimer disease and cerebrovascular pathology: an update. *J Neural Transm (Vienna), 109*(5-6), 813-836. doi:10.1007/s007020200068

Jonsson, H., Helgadottir, G. P., Aspelund, T., Eiriksdottir, G., Sigurdsson, S., Ingvarsson, T., . . . Gudnason, V. (2009). Hand osteoarthritis in older women is associated with carotid and coronary

atherosclerosis: the AGES Reykjavik study. *Ann Rheum Dis, 68*(11), 1696-1700. doi:10.1136/ard.2008.096289
Karch, C. M., & Goate, A. M. (2015). Alzheimer's disease risk genes and mechanisms of disease pathogenesis. *Biol Psychiatry, 77*(1), 43-51. doi:10.1016/j.biopsych.2014.05.006
Khan, T. A., Shah, T., Prieto, D., Zhang, W., Price, J., Fowkes, G. R., . . . Casas, J. P. (2013). Apolipoprotein E genotype, cardiovascular biomarkers and risk of stroke: systematic review and meta-analysis of 14,015 stroke cases and pooled analysis of primary biomarker data from up to 60,883 individuals. *Int J Epidemiol, 42*(2), 475-492. doi:10.1093/ije/dyt034
Klein, R., Meuer, S. M., Moss, S. E., Klein, B. E., Neider, M. W., & Reinke, J. (2004). Detection of age-related macular degeneration using a nonmydriatic digital camera and a standard film fundus camera. *Arch Ophthalmol, 122*(11), 1642-1646. doi:10.1001/archopht.122.11.1642
Knudtson, M. D., Lee, K. E., Hubbard, L. D., Wong, T. Y., Klein, R., & Klein, B. E. (2003). Revised formulas for summarizing retinal vessel diameters. *Curr Eye Res, 27*(3), 143-149.
Liu, G., Yao, L., Liu, J., Jiang, Y., Ma, G., Genetic, . . . Li, K. (2014). Cardiovascular disease contributes to Alzheimer's disease: evidence from large-scale genome-wide association studies. *Neurobiol Aging, 35*(4), 786-792. doi:10.1016/j.neurobiolaging.2013.10.084
Liu, W., Zhang, Y., Yu, C. M., Ji, Q. W., Cai, M., Zhao, Y. X., & Zhou, Y. J. (2015). Current understanding of coronary artery calcification. *J Geriatr Cardiol, 12*(6), 668-675. doi:10.11909/j.issn.1671-5411.2015.06.012
Lloyd-Jones, D. M., Hong, Y., Labarthe, D., Mozaffarian, D., Appel, L. J., Van Horn, L., . . . Statistics, C. (2010). Defining and setting national goals for cardiovascular health promotion and disease reduction: the American Heart Association's strategic Impact Goal through 2020 and beyond. *Circulation, 121*(4), 586-613. doi:10.1161/CIRCULATIONAHA.109.192703
Meier, I. B., Gu, Y., Guzaman, V. A., Wiegman, A. F., Schupf, N., Manly, J. J., . . . Brickman, A. M. (2014). Lobar microbleeds are associated with a decline in executive functioning in older adults. *Cerebrovasc Dis, 38*(5), 377-383. doi:10.1159/000368998
Michaelson, D. M. (2014). APOE epsilon4: the most prevalent yet understudied risk factor for Alzheimer's disease. *Alzheimers Dement, 10*(6), 861-868. doi:10.1016/j.jalz.2014.06.015
Middleton, L. E., & Yaffe, K. (2009). Promising strategies for the prevention of dementia. *Arch Neurol, 66*(10), 1210-1215. doi:10.1001/archneurol.2009.201
Mormino, E. C., Sperling, R. A., Holmes, A. J., Buckner, R. L., De Jager, P. L., Smoller, J. W., . . . Alzheimer's Disease Neuroimaging, I. (2016). Polygenic risk of Alzheimer disease is associated with early- and late-life processes. *Neurology, 87*(5), 481-488. doi:10.1212/WNL.0000000000002922
Mucke, L. (2009). Neuroscience: Alzheimer's disease. *Nature, 461*(7266), 895-897. doi:10.1038/461895a
Nagasawa, J., Kiyozaka, T., & Ikeda, K. (2014). Prevalence and clinicoradiological analyses of patients with Alzheimer disease coexisting multiple microbleeds. *J Stroke Cerebrovasc Dis, 23*(9), 2444-2449. doi:10.1016/j.jstrokecerebrovasdis.2014.05.036
Naj, A. C., Jun, G., Beecham, G. W., Wang, L. S., Vardarajan, B. N., Buros, J., . . . Schellenberg, G. D. (2011). Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat Genet, 43*(5), 436-441. doi:10.1038/ng.801
Newman, A. B., Fitzpatrick, A. L., Lopez, O., Jackson, S., Lyketsos, C., Jagust, W., . . . Kuller, L. H. (2005). Dementia and Alzheimer's disease incidence in relationship to cardiovascular disease in the Cardiovascular Health Study cohort. *J Am Geriatr Soc, 53*(7), 1101-1107. doi:10.1111/j.1532-5415.2005.53360.x
Pearl, J. (2001). *Direct and Indirect Effects.* Retrieved from San Francisco, CA: http://ftp.cs.ucla.edu/pub/stat_ser/R273-U.pdf
Pfeifer, L. A., White, L. R., Ross, G. W., Petrovitch, H., & Launer, L. J. (2002). Cerebral amyloid angiopathy and cognitive function: the HAAS autopsy study. *Neurology, 58*(11), 1629-1634.

Poels, M. M., Ikram, M. A., van der Lugt, A., Hofman, A., Niessen, W. J., Krestin, G. P., . . . Vernooij, M. W. (2012). Cerebral microbleeds are associated with worse cognitive function: the Rotterdam Scan Study. *Neurology, 78*(5), 326-333. doi:10.1212/WNL.0b013e3182452928

Prins, N. D., van Dijk, E. J., den Heijer, T., Vermeer, S. E., Koudstaal, P. J., Oudkerk, M., . . . Breteler, M. M. (2004). Cerebral white matter lesions and the risk of dementia. *Arch Neurol, 61*(10), 1531-1534. doi:10.1001/archneur.61.10.1531

Qiu, C., Cotch, M. F., Sigurdsson, S., Garcia, M., Klein, R., Jonasson, F., . . . Launer, L. J. (2008). Retinal and cerebral microvascular signs and diabetes: the age, gene/environment susceptibility-Reykjavik study. *Diabetes, 57*(6), 1645-1650. doi:10.2337/db07-1455

Reis, J. P., Launer, L. J., Terry, J. G., Loria, C. M., Zeki Al Hazzouri, A., Sidney, S., . . . Carr, J. J. (2013). Subclinical atherosclerotic calcification and cognitive functioning in middle-aged adults: the CARDIA study. *Atherosclerosis, 231*(1), 72-77. doi:10.1016/j.atherosclerosis.2013.08.038

Robins, J. M., & Greenland, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology, 3*(2), 143-155.

Saczynski, J. S., Jonsdottir, M. K., Sigurdsson, S., Eiriksdottir, G., Jonsson, P. V., Garcia, M. E., . . . Launer, L. J. (2008). White matter lesions and cognitive performance: the role of cognitively complex leisure activity. *J Gerontol A Biol Sci Med Sci, 63*(8), 848-854.

Saczynski, J. S., Sigurdsson, S., Jonsdottir, M. K., Eiriksdottir, G., Jonsson, P. V., Garcia, M. E., . . . Launer, L. J. (2009). Cerebral infarcts and cognitive performance: importance of location and number of infarcts. *Stroke, 40*(3), 677-682. doi:10.1161/STROKEAHA.108.530212

Sajeev, G. (2015). *Mediation Analysis in Understanding Mechanism of Alzheimer's Disease Risk.* (Doctoral), Harvard T.H. Chan School of Public Health, Boston. Retrieved from https://dash.harvard.edu/handle/1/16121154

Schellenberg, G. D., & Montine, T. J. (2012). The genetics and neuropathology of Alzheimer's disease. *Acta Neuropathol, 124*(3), 305-323. doi:10.1007/s00401-012-0996-2

Schilling, S., DeStefano, A. L., Sachdev, P. S., Choi, S. H., Mather, K. A., DeCarli, C. D., . . . Debette, S. (2013). APOE genotype and MRI markers of cerebrovascular disease: systematic review and meta-analysis. *Neurology, 81*(3), 292-300. doi:10.1212/WNL.0b013e31829bfda4

Schneider, J. A., Arvanitakis, Z., Bang, W., & Bennett, D. A. (2007). Mixed brain pathologies account for most dementia cases in community-dwelling older persons. *Neurology, 69*(24), 2197-2204. doi:10.1212/01.wnl.0000271090.28148.24

Schneider, J. A., Arvanitakis, Z., Leurgans, S. E., & Bennett, D. A. (2009). The neuropathology of probable Alzheimer disease and mild cognitive impairment. *Ann Neurol, 66*(2), 200-208. doi:10.1002/ana.21706

Schneider, J. A., & Bennett, D. A. (2010). Where vascular meets neurodegenerative disease. *Stroke, 41*(10 Suppl), S144-146. doi:10.1161/STROKEAHA.110.598326

Schneider, J. A., Wilson, R. S., Bienias, J. L., Evans, D. A., & Bennett, D. A. (2004). Cerebral infarctions and the likelihood of dementia from Alzheimer disease pathology. *Neurology, 62*(7), 1148-1155.

Silvestrini, M., Gobbi, B., Pasqualetti, P., Bartolini, M., Baruffaldi, R., Lanciotti, C., . . . Vernieri, F. (2009). Carotid atherosclerosis and cognitive decline in patients with Alzheimer's disease. *Neurobiol Aging, 30*(8), 1177-1183. doi:10.1016/j.neurobiolaging.2007.11.008

Smith, E. E., Nandigam, K. R., Chen, Y. W., Jeng, J., Salat, D., Halpin, A., . . . Greenberg, S. M. (2010). MRI markers of small vessel disease in lobar and deep hemispheric intracerebral hemorrhage. *Stroke, 41*(9), 1933-1938. doi:10.1161/STROKEAHA.110.579078

Staals, J., Booth, T., Morris, Z., Bastin, M. E., Gow, A. J., Corley, J., . . . Wardlaw, J. M. (2015). Total MRI load of cerebral small vessel disease and cognitive ability in older people. *Neurobiol Aging, 36*(10), 2806-2811. doi:10.1016/j.neurobiolaging.2015.06.024

Sturlaugsdottir, R., Aspelund, T., Bjornsdottir, G., Sigurdsson, S., Eiriksdottir, G., Imai, C. M., . . . Gudnason, V. (2015). Carotid atherosclerosis and cardiovascular health metrics in old subjects from the AGES-Reykjavik study. *Atherosclerosis, 242*(1), 65-70. doi:10.1016/j.atherosclerosis.2015.06.043

Sveinbjornsdottir, S., Sigurdsson, S., Aspelund, T., Kjartansson, O., Eiriksdottir, G., Valtysdottir, B., . . . Launer, L. J. (2008). Cerebral microbleeds in the population based AGES-Reykjavik study: prevalence and location. *J Neurol Neurosurg Psychiatry, 79*(9), 1002-1006. doi:10.1136/jnnp.2007.121913

Tai, L. M., Thomas, R., Marottoli, F. M., Koster, K. P., Kanekiyo, T., Morris, A. W., & Bu, G. (2016). The role of APOE in cerebrovascular dysfunction. *Acta Neuropathol, 131*(5), 709-723. doi:10.1007/s00401-016-1547-z

Toledo, J. B., Arnold, S. E., Raible, K., Brettschneider, J., Xie, S. X., Grossman, M., . . . Trojanowski, J. Q. (2013). Contribution of cerebrovascular disease in autopsy confirmed neurodegenerative disease cases in the National Alzheimer's Coordinating Centre. *Brain, 136*(Pt 9), 2697-2706. doi:10.1093/brain/awt188

Traylor, M., Adib-Samii, P., Harold, D., Alzheimer's Disease Neuroimaging, I., International Stroke Genetics Consortium, U. K. Y. L. S. D. N. A. r., Dichgans, M., . . . International Genomics of Alzheimer's Project, i. (2016). Shared genetic contribution to Ischaemic Stroke and Alzheimer's Disease. *Ann Neurol*. doi:10.1002/ana.24621

Valeri, L., & Vanderweele, T. J. (2013). Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. *Psychol Methods, 18*(2), 137-150. doi:10.1037/a0031034

van Oijen, M., de Jong, F. J., Witteman, J. C., Hofman, A., Koudstaal, P. J., & Breteler, M. M. (2007). Atherosclerosis and risk for dementia. *Ann Neurol, 61*(5), 403-410. doi:10.1002/ana.21073

Vanderweele, T. J. (2015). *Explanation in causal inference : methods for mediation and interaction.* . New York: Oxford University Press.

VanderWeele, T. J. (2016). Mediation Analysis: A Practitioner's Guide. *Annu Rev Public Health, 37*, 17-32. doi:10.1146/annurev-publhealth-032315-021402

Vernooij, M. W., van der Lugt, A., Ikram, M. A., Wielopolski, P. A., Niessen, W. J., Hofman, A., . . . Breteler, M. M. (2008). Prevalence and risk factors of cerebral microbleeds: the Rotterdam Scan Study. *Neurology, 70*(14), 1208-1214. doi:10.1212/01.wnl.0000307750.41970.d9

Wendell, C. R., Waldstein, S. R., Ferrucci, L., O'Brien, R. J., Strait, J. B., & Zonderman, A. B. (2012). Carotid atherosclerosis and prospective risk of dementia. *Stroke, 43*(12), 3319-3324. doi:10.1161/STROKEAHA.112.672527

Weuve, J. M., M. B.; Blacker, D. (2017). The AlzRisk Database. Alzheimer Research Forum. Available at: http://www.alzforum.org. Retrieved 08/16/2017

Wisdom, N. M., Callahan, J. L., & Hawkins, K. A. (2011). The effects of apolipoprotein E on non-impaired cognitive functioning: a meta-analysis. *Neurobiol Aging, 32*(1), 63-74. doi:10.1016/j.neurobiolaging.2009.02.003

Yates, P. A., Villemagne, V. L., Ellis, K. A., Desmond, P. M., Masters, C. L., & Rowe, C. C. (2014). Cerebral microbleeds: a review of clinical, genetic, and neuroimaging associations. *Front Neurol, 4*, 205. doi:10.3389/fneur.2013.00205

Table 2.1: Descriptive statistics of demographic and clinical characteristics

Characteristics	All subjects	Subjects with genotype data
	(N=5161)	(N=2907)
Demographic		
Age at AGES I (years), *mean (SD)*	76.73 (5.83)	76.20 (5.43)
Sex		
Female, *N (%)*	3022 (58.6)	1706 (58.7)
Education		
Secondary, *N (%)*	2406 (49.9)	1441 (49.7)
College, *N (%)*	755 (15.7)	451 (15.6)
University, *N (%)*	547 (11.4)	334 (11.5)
Vascular Pathologies, Baseline		
Lobar cerebral microbleeds		
Count>=2, *N(%)*	110 (2.1)	69 (2.7)
White matter lesion load, *median(Q1, Q3)*	1.91 (0.51, 5.64)	1.92 (0.50, 5.59)
Central retinal venular equivalent, *mean(SD)*	202.19 (19.56)	202.14 (19.50)
Carotid intima-media thickness, *median(Q1,Q3)*	0.97 (0.88, 1.06)	0.97 (0.88, 1.06)
Coronary calcification score, *median(Q1, Q3)*	271.23 (43.61, 898.78)	253.52 (38.94, 841.53)
Other Covariates, Baseline		
Midlife physical activity		
Intermediate, *N(%)*	2166 (46.6)	1327 (47.5)
Poor, *N(%)*	909 (19.5)	524 (18.8)
Diet quality		
Intermediate, *N(%)*	4011 (84.6)	2418 (84.8)
Poor, *N(%)*	354 (7.5)	205 (7.2)
Smoking		
Ever, *N(%)*	2111 (43.9)	1303 (44.8)
Current, *N(%)*	593 (12.3)	372 (12.8)
Diabetes, *N(%)*	640 (12.4)	324 (11.2)
Hypertension		
Prehypertension, *N(%)*	758 (14.8)	445 (15.3)
Hypertension, *N(%)*	4112 (80.3)	2318 (79.8)
LDL level >=130 mg/dL, *N(%)*	2830 (54.9)	1643 (56.6)
BMI>=30, *N(%)*	1139 (22.3)	642 (22.1)

Subjects with genotype data had lower coronary calcification score (P=0.01) than other subjects. No significant difference was observed in the distribution of any other baseline characteristic listed in the Table between subjects with and without GWAS genotype data available.

Table 2.2: Phenotypic associations between vascular pathologies and late-life cognitive function

	Memory				Global Cognition			
	Unadjusted		Adjusted		Unadjusted		Adjusted	
	Beta (S.E.)	**P**	**Beta (S.E.)**	**P**	**Beta (S.E.)**	**P**	**Beta (S.E.)**	**P**
Lobar CMB	-0.3684 (0.0973)	1.5E-04*	-0.1868 (0.0857)	0.03*	-0.4434 (0.0997)	8.9E-06*	-0.2679 (0.0826)	0.001*
WML	-0.1925 (0.0362)	1.1E-07*	-0.0696 (0.0328)	0.03*	-0.2440 (0.0372)	6.0E-11*	-0.0596 (0.0318)	0.06
RVD	0.0020 (0.0008)	0.01*	-0.0005 (0.0007)	0.53	0.0027 (0.0008)	0.001*	0.0004 (0.0007)	0.56
CIMT	-1.1099 (0.1047)	6.3E-26*	-0.1066 (0.1015)	0.29	-1.0042 (0.1089)	4.5E-20	0.0731 (0.0987)	0.46
CAC	-0.0858 (0.0059)	1.5E-46*	-0.0154 (0.0061)	0.01*	-0.0872 (0.0061)	2.4E-45	-0.0232 (0.0059)	7.8E-05*

Lobar CMB: Multiple lobar cerebral microbleeds; count >=2 vs. 0 or 1.
WML: White matter lesion load; highest quartile vs. other three quartiles of the total volume of white matter lesions.
RVD: Retinal venular diameter; represented by central retinal venular equivalent, continuous.
CIMT: Carotid intima-media thickness; log-transformed, continuous.
CAC: Coronary artery calcification score; log-transformed, continuous.

All multivariate models adjusted for age, sex, education, diabetes, hypertension, high LDL level, obesity, physical activity, diet quality, and smoking status.

Asterisk indicates significance at P<0.05.

Table 2.3: Associations between polygenic risk scores for AD dementia and each target phenotype

PRS for AD dementia			Target Phenotypes						
			Cognitive Outcomes		Vascular Markers				
Genomic region	**P-threshold**	**nSNP**	**Memory** (n=2752)	**Global Cognition** (n=2582)	**Lobar CMB** (n=2562)	**WML** (n=2559)	**RVD** (n=2682)	**CIMT** (n=2777)	**CAC** (n=2869)
GW-ADPRS	$P_T < 0.0001$	190	R^2=0.0022 P=0.006**	R^2=0.0015 P=0.02*	R^2=0.0030 P=0.006*	R^2=0.0012 P=0.07	R^2=0.0004 P=0.29	R^2<0.0001 P=0.71	R^2=0.0014 P=0.03*
	$P_T < 0.001$	1342	R^2=0.0012 P=0.04*	R^2=0.0006 P=0.14	R^2=0.0029 P=0.007*	R^2=0.0011 P=0.09	R^2<0.0001 P=0.90	R^2=0.0001 P=0.61	R^2=0.0004 P=0.24
	$P_T < 0.01$	8918	R^2=0.0029 P=0.001**	R^2=0.0012 P=0.04*	R^2=0.0018 P=0.03*	R^2=0.0018 P=0.03*	R^2=0.0002 P=0.49	R^2<0.0001 P=0.99	R^2=0.0005 P=0.19
19q13-ADPRS	$P_T < 0.0001$	40	R^2=0.0015 P=0.02*	R^2=0.0013 P=0.04*	R^2=0.0038 P=0.002**	R^2=0.0014 P=0.05	R^2=0.0003 P=0.40	R^2=0.0002 P=0.46	R^2=0.0012 P=0.04*
	$P_T < 0.001$	54	R^2=0.0015 P=0.02*	R^2=0.0013 P=0.04*	R^2=0.0035 P=0.003**	R^2=0.0018 P=0.03*	R^2=0.0002 P=0.48	R^2=0.0002 P=0.48	R^2=0.0015 P=0.02*
	$P_T < 0.01$	76	R^2=0.0012 P=0.04*	R^2=0.0011 P=0.06	R^2=0.0040 P=0.002**	R^2=0.0019 P=0.02*	R^2=0.0002 P=0.46	R^2=0.0003 P=0.30	R^2=0.0015 P=0.02*
Non-19q13-ADPRS	$P_T < 0.0001$	150	R^2=0.0008 P=0.10	R^2=0.0002 P=0.39	R^2<0.0001 P=0.94	R^2<0.0001 P=0.82	R^2=0.0002 P=0.50	R^2=0.0001 P=0.50	R^2=0.0001 P=0.49
	$P_T < 0.001$	1288	R^2=0.0002 P=0.45	R^2<0.0001 P=0.81	R^2=0.0004 P=0.30	R^2=0.0001 P=0.68	R^2=0.0001 P=0.67	R^2<0.0001 P=0.94	R^2<0.0001 P=0.71
	$P_T < 0.01$	8842	R^2=0.0020 P=0.008**	R^2=0.0006 P=0.15	R^2=0.0005 P=0.27	R^2=0.0008 P=0.14	R^2=0.0004 P=0.31	R^2<0.0001 P=0.70	R^2=0.0001 P=0.59

P-threshold (P_T): the P-value threshold used in the training dataset to select SNPs for calculating the PRS for AD dementia.
nSNP: different number of independent SNPs included for calculating the PRS for AD dementia, which is determined by the selection of P_T.
Lobar CMB: lobar cerebral microbleeds (count >=2 vs. 0 or 1).
WML: total brain white matter lesion load (highest quartile vs. other three quartiles of the total volume of white matter lesions).
RVD: the average retinal venular diameter, represented by central retinal venular equivalent (continuous).
CIMT: mean of carotid intima-media thickness (log-transformed, continuous).
CAC: coronary artery calcification score (log-transformed, continuous).

R^2: squared semi-partial correlation, the proportion of variance in the target phenotype explained by the PRS for AD dementia.
P: the P-value of the test for association between the PRS and the target phenotype, before multiple testing correction.
An asterisk indicates significance at P-value <0.05.
A double asterisk indicates Boferroni-corrected P-value < 0.05.
Associations were tested using linear (for continuous phenotype) or logistic (for binary phenotype) regression models with age and sex as covariates.
The missing data status of each vascular marker was associated neither with memory / global cognition scores nor with any of the PRS (all P > 0.20).

Table 2.4: Total, direct and indirect effects of PRS for AD dementia on late-life cognitive function mediated by vascular pathologies

			Memory				Global Cognition			
A	**M**	**N**	**Total effect (95% BCCI)**	**Direct effect (95% BCCI)**	**Indirect effect (95% BCCI)**	**PM**	**Total effect (95% BCCI)**	**Direct effect (95% BCCI)**	**Indirect effect (95% BCCI)**	**PM**
GW-ADPRS	CMB	2308	-.0743 (-.1165, -.0249)	-.0719 (-.1145, -.0231)	-.0024 (-.0058, -.0005)	3.2% *	-.0377 (-.0808, .0079)	-.0336 (-.0763, .0128)	-.0041 (-.0091, -.0015)	10.8% *
	WML	2307	-.0783 (-.1242, -.0330)	-.0762 (-.1228, -.0317)	-.0021 (-.0070, -.00004)	2.7% *	-.0409 (-.0872, .0107)	-.0393 (-.0847, .0126)	-.0016 (-.0062, -.0003)	4.0% *
	CAC	2577	-.0685 (-.1118, -.0232)	-.0674 (-.1102, -.0220)	-.0011 (-.0057, .00005)	1.6%	-.0483 (-.0926, -.0032)	-.0462 (-.0919, -.0014)	-.0021 (-.0060, -.0001)	4.4% *
19q13-ADPRS	CMB	2308	-.0376 (-.0871, .0130)	-.0348 (-.0848, .0167)	-.0028 (-.0075, -.0005)	7.4% *	-.0363 (-.0902, .0090)	-.0322 (-.0792, .0151)	-.0041 (-.0128, -.0013)	11.3% *
	WML	2307	-.0434 (-.0912, .0073)	-.0410 (-.0878, .0098)	-.0024 (-.0080, -.0004)	5.6% *	-.0400 (-.0910, .0079)	-.0381 (-.0875, .0122)	-.0019 (-.0069, -.0002)	4.8% *
	CAC	2577	-.0481 (-.0958, -.0014)	-.0463 (-.0932, .0018)	-.0018 (-.0054, .0004)	3.6%	-.0478 (-.0996, -.0025)	-.0454 (-.0960, -.0004)	-.0024 (-.0063, -.00005)	5.0% *
Non-19q13-	CMB	2308	-.0679 (-.1176, -.0219)	-.0665 (-.1160, -.0198)	-.0014 (-.0046, -.0001)	2.1% *	-.0279 (-.0750, .0163)	-.0258 (-.0735, .0194)	-.0021 (-.0063, -.00005)	7.5% *
	WML	2307	-.0703 (-.1145, -.0223)	-.0689 (-.1138, -.0207)	-.0014 (-.0050, .0003)	2.0%	-.0293 (-.0745, .0165)	-.0282 (-.0732, .0178)	-.0011 (-.0050, .0003)	3.7%
	CAC	2577	-.0576 (-.1024, -.0129)	-.0571 (-.1012, -.0120)	-.0005 (-.0037, .0005)	0.9%	-.0314 (-.0779, .0152)	-.0307 (-.0767, .0151)	-.0007 (-.0041, .0006)	2.2%

Exposure (A): GW-ADPRS, 19q13-ADPRS, and non-19q13-ADPRS (the 75th percentile vs. the 25th percentile)
Mediator (M): CMB (lobar cerebral microbleeds; >=2 vs. 0 or 1), WML (total brain white matter lesion load; highest quartile vs. other three quartiles of the total volume of white matter lesions), or CAC (coronary artery calcification score; log-transformed, continuous)
Outcome: z-standardized memory composite score (left panel) or z-standardized global cognition composite score (right panel)
Values for total, direct and indirect effects indicate changes in each outcome.
BCCI: bias-corrected confidence interval
PM: proportion mediated=indirect effect beta coefficient/ total effect beta coefficient
Models for the effects of GW-PRS adjusted for age, sex, smoking status, midlife physical activity, and diet quality.
Models for the effects of 19q13-ADPRS adjusted for age, sex, smoking status, midlife physical activity, diet quality, and non-19q13-ADPRS
Models for the effects of non-19q13-ADPRS adjusted for age, sex, smoking status, midlife physical activity, diet quality, and 19q13-ADPRS

Table 2.5: Phenotypic associations between vascular pathologies and late-life cognitive function, additionally adjusting for shared genetic factors

	Memory		**Global Cognition**	
	Beta (S.E.)	**P**	**Beta (S.E.)**	**P**
Lobar CMB	-0.2269 (0.1070)	0.03*	-0.3339 (0.1013)	0.001*
WML	-0.1009 (0.0424)	0.02*	-0.0788 (0.0405)	0.05
RVD	-0.0005 (0.0009)	0.58	0.0001 (0.0009)	0.88
CIMT	0.0017 (0.1292)	0.99	0.1317 (0.1243)	0.29
CAC	-0.0118 (0.0077)	0.13	-0.0152 (0.0075)	0.04*

Lobar CMB: Multiple lobar cerebral microbleeds; count >=2 vs. 0 or 1.
WML: White matter lesion load; highest quartile vs. other three quartiles of the total volume of white matter lesions.
RVD: Retinal venular diameter; represented by central retinal venular equivalent, continuous.
CIMT: Carotid intima-media thickness; log-transformed, continuous.
CAC: Coronary artery calcification score; log-transformed, continuous.

All models adjusted for age, sex, education, diabetes, hypertension, high LDL level, obesity, physical activity, diet quality, smoking status, 19q13-ADPRS$_{PT = 0.001}$, and non-19q13-ADPRS$_{PT = 0.01}$.

Asterisk indicates significance at P-value <0.05.

Figure 2.1: The study samples used in the analyses

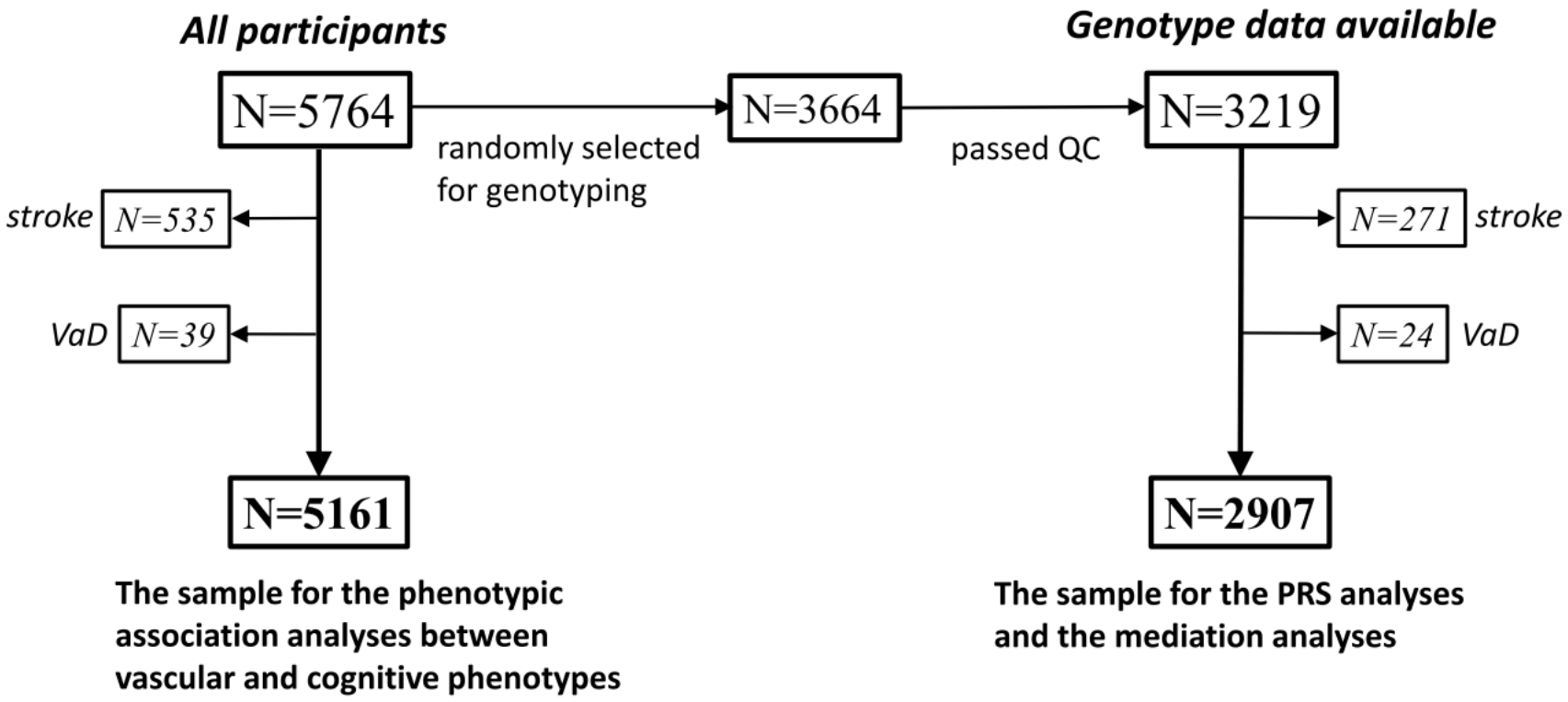
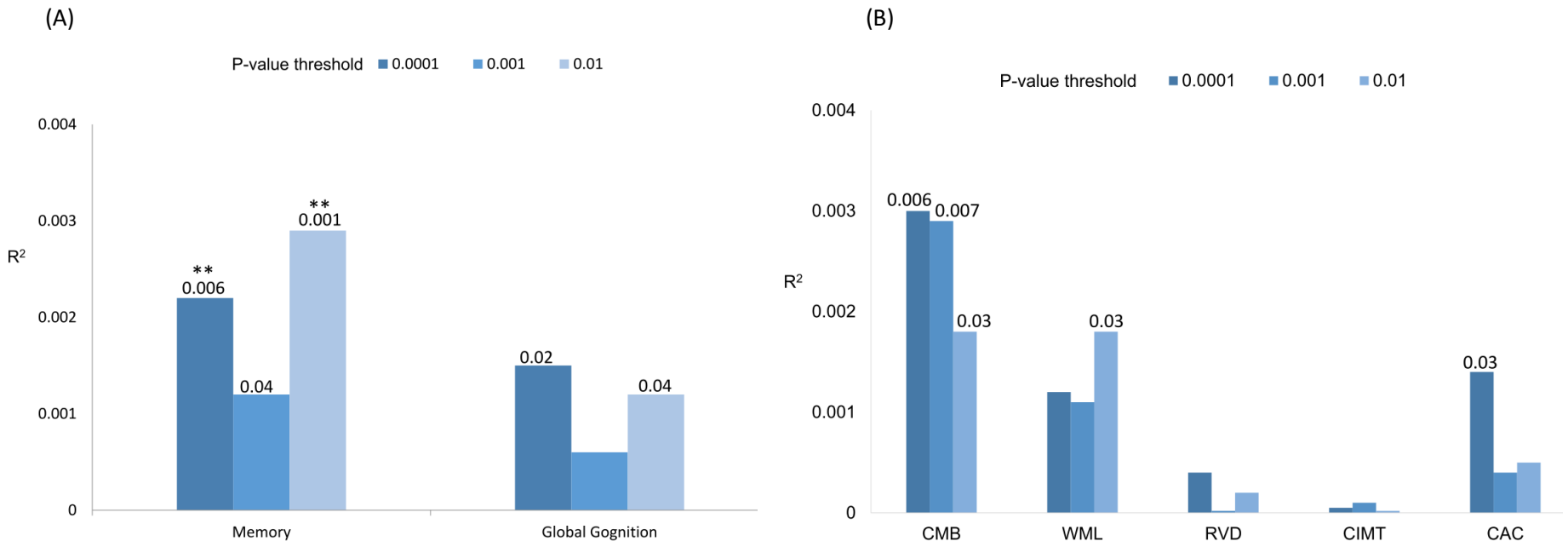
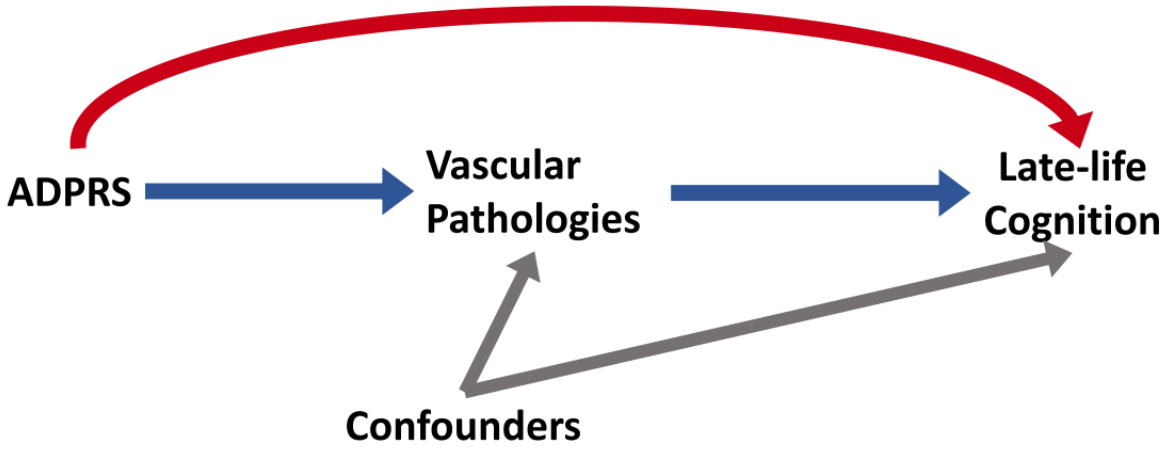


Figure 2.2: Pair-wise polygenic association analyses between GW-ADPRS and (A) cognition scores, (B) vascular pathologies



We derived genome-wide PRS for AD dementia using ADGC GWAS as the discovery sample with three different P-value thresholds (P_T used to select training set SNPs: 0.0001, 0.001, and 0.01) and apply them to (A) Z-score of the composite memory or global cognition score; and (B) each of the markers of vascular pathologies. Age and sex were included as covariates in the association analyses.
Each pair is shown on the x-axis and the proportion of variance explained for each phenotype (estimated via partial correlation R^2) on the y-axis.
Unadjusted P-values are shown on the top of the bars if < 0.05.
Double asterisks indicate Bonferroni-corrected P-value < 0.05.

Figure 2.3: Directed acyclic graph (DAG) for the relationship between PRS for AD dementia, vascular pathologies, and late-life cognition



Exposure: ADPRS, PRS for AD dementia

Mediator: vascular pathologies, including CMB, WML, and CAC

Outcome: late-life memory and global cognition scores

The path traced by the red arrow represents the direct effect of the ADPRS on cognition, while the path traced by the blue arrows represents the indirect (mediated) effect through vascular pathologies.

Genetic overlap between vascular pathologies and Alzheimer's dementia and the potential causal mechanisms

Supplementary Information

SUPPLEMENTARY METHODS

Sensitivity Analyses of Unmeasured Confounding

The counterfactual-based mediation analysis assumes no unmeasured confounding for the (1) exposure-mediator, (2) exposure-outcome, and (3) mediator-outcome relationships (Vanderweele, 2015). In our mediation analyses with genetic risk scores as the exposures, assumptions (1) and (2) were probably plausible. However, the assumption of no unmeasured confounding might be less plausible for the (3) mediator-outcome relationship, and the effect estimates would probably be biased.

In order to evaluate the robustness of the mediation analyses to unmeasured confounding between mediator and outcome, we conducted sensitivity analyses to calculate how much the estimated mediated effects would be expected to change under different degrees of mediator-outcome confounding. Specifically, a mediator-outcome confounder leads to a correlation between the residuals in the two regression models predicting mediator and outcome.

We generalize the regression models for mediation analysis as:

$$M_i = \alpha_1 + \beta_1 A_i + \xi_1^A X_i + e_{i1}$$

$$Y_i = \alpha_2 + \beta_2 A_i + \gamma M_i + \kappa A_i M_i + \xi_2^A X_i + e_{i2}$$

A = exposure, M = mediator, Y = outcome, X= covariates, e = residual

The residual accounts for the variability in the dependent variable that cannot be explained by the linear relationship with the independent variables. The sensitivity parameter, denoted by ρ, is the correlation between e_{i1} and e_{i2}.
Under the assumption of no unmeasured mediator-outcome confounding, ρ is equal to zero. The magnitude of ρ represents the departure from the assumption of no unmeasured mediator-outcome confounding.

We explored the sensitivity of mediation results to a hypothetical confounder by systematically increasing the correlation between the residuals in the regression models predicting mediator and outcome (Imai et al., 2010; Imai, 2010). We used an R program developed by Imai et al. to assess how large a confounder effect on the mediator-outcome relation must be to invalidate the conclusion for each mediation analysis.

SUPPLEMENTARY RESULTS

Sensitivity Analyses of Unmeasured Confounding

Results of sensitivity analyses of unmeasured confounding are presented in Figure S2.1. When rho<0, which means there is unmeasured confounding (or residual confounding) associated with better cognition and less severe vascular pathology or unmeasured confounding associated with poorer cognition and more severe vascular pathology, our estimated PMs would underestimate the true mediation effects of vascular pathologies. When rho>0, which means there is unmeasured confounding associated with better cognition and more severe vascular pathology, or unmeasured confounding associated with poorer cognition and less severe vascular pathology, our estimated PMs would overestimate the true mediation. The values of rho at which proportion mediated equals to 0 for the mediation models are listed in Table S2.2. The value of rho at which proportion mediated equals to 0 represents the degree of confounding that would change the conclusion on the mediation effect by vascular pathologies.

SUPPLEMENTARY TABLES

Table S2.1: Total, direct and indirect effects of PRS for AD dementia on late-life cognitive function mediated by vascular pathologies

			Memory				Global Cognition			
A	**M**	**N**	**Total effect (95% BCCI)**	**Direct effect (95% BCCI)**	**Indirect effect (95% BCCI)**	**PM**	**Total effect (95% BCCI)**	**Direct effect (95% BCCI)**	**Indirect effect (95% BCCI)**	**PM**
GW-ADPRS	CMB	2308	-.1434 (-.2239, -.0469)	-.1387 (-.2207, -.0446)	-.0047 (-.0119, -.0011)	3.3% *	-.0708 (-.1515, .0148)	-.0628 (-.1428, .0240)	-.0080 (-.0201, -.0030)	11.4% *
	WML	2307	-.1510 (-.2394, -.0635)	-.1470 (-.2367, -.0612)	-.0040 (-.0135, -.00007)	2.7% *	-.0766 (-.1632, .0201)	-.0735 (-.1585, .0235)	-.0030 (-.0117, -.0006)	4.0% *
	CAC	2577	-.1320 (-.2155, -.0447)	-.1300 (-.2125, -.0424)	-.0021 (-.0109, .00009)	1.6%	-.0904 (-.1732, -.0059)	-.0864 (-.1718, -.0026)	-.0040 (-.0113, -.0002)	4.4% *
19q13-ADPRS	CMB	2308	-.0689 (-.1590, .0234)	-.0635 (-.1546, .0305)	-.0054 (-.0157, -.0010)	7.9% *	-.0666 (-.1645, .0151)	-.0586 (-.1443, .0275)	-.0080 (-.0271, -.0026)	12.0% *
	WML	2307	-.0791 (-.1661, .0134)	-.0747 (-.1601, .0179)	-.0044 (-.0149, -.0007)	5.6% *	-.0729 (-.1658, .0145)	-.0694 (-.1596, .0223)	-.0035 (-.0128, -.0004)	4.8% *
	CAC	2577	-.0876 (-.1746, -.0025)	-.0844 (-.1700, .0033)	-.0032 (-.0098, .0007)	3.6%	-.0871 (-.1815, -.0046)	-.0827 (-.1750, -.0008)	-.0043 (-.0114, -.0001)	5.0% *
Non-19q13-	CMB	2308	-.1337 (-.2317, -.0431)	-.1309 (-.2285, -.0391)	-.0028 (-.0093, -.0002)	2.1% *	-.0549 (-.1422, .0368)	-.0507 (-.1364, .0428)	-.0041(-.0138,-9.6E-06)	7.5% *
	WML	2307	-.1384 (-.2256, -.0440)	-.1357 (-.2242, -.0407)	-.0027 (-.0098, .0005)	2.0%	-.0576 (-.1514, .0302)	-.0555 (-.1492, .0325)	-.0021 (-.0093, .0007)	3.7%
	CAC	2577	-.1135 (-.2017, -.0253)	-.1125 (-.1993, -.0237)	-.0010 (-.0073, .0010)	0.9%	-.0618 (-.1534, .0299)	-.0605 (-.1510, .0298)	-.0013 (-.0080, .0012)	2.2%

Exposure (A): GW-ADPRS, 19q13-ADPRS, and non-19q13-ADPRS (the 90th percentile vs. the 10th percentile)
Mediator (M): CMB (lobar cerebral microbleeds; >=2 vs. 0 or 1), WML (total brain white matter lesion load; highest quartile vs. other three quartiles of the total volume of white matter lesions), or CAC (coronary artery calcification score; log-transformed, continuous)
Outcome: z-standardized memory composite score (left panel) or z-standardized global cognition composite score (right panel)
Values for total, direct and indirect effects indicate changes in each outcome.
BCCI: bias-corrected confidence interval
PM: proportion mediated=indirect effect beta coefficient/ total effect beta coefficient
Models for the effects of GW-PRS adjusted for age, sex, smoking status, midlife physical activity, and diet quality.
Models for the effects of 19q13-ADPRS adjusted for age, sex, smoking status, midlife physical activity, diet quality, and non-19q13-ADPRS
Models for the effects of non-19q13-ADPRS adjusted for age, sex, smoking status, midlife physical activity, diet quality, and 19q13-ADP

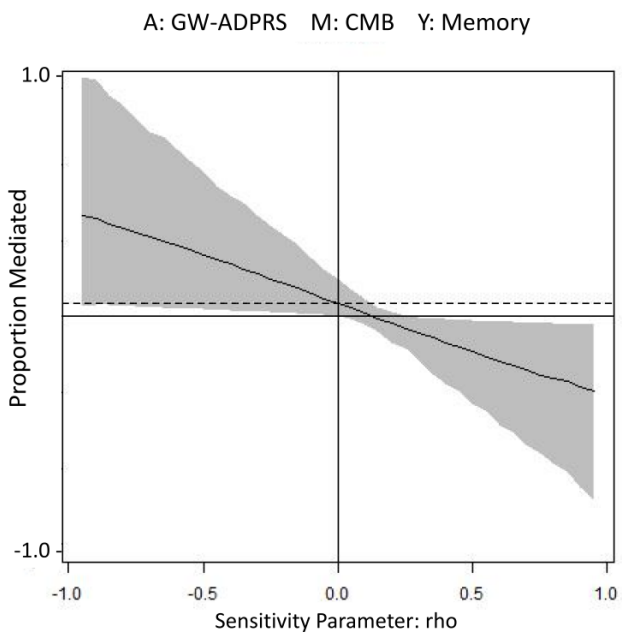
Table S2.2: The values of rho at which proportion mediated equals to 0 for mediation models

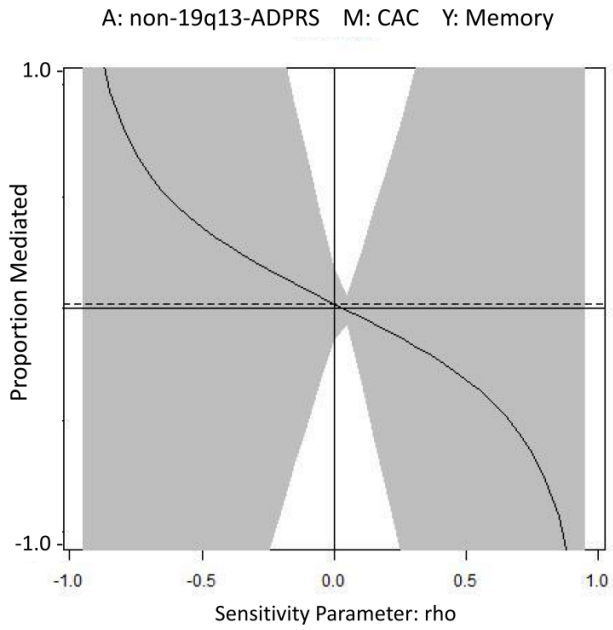
Exposure	Mediator	Outcome	Rho at PM=0
GW-ADPRS	CMB	Memory	0.13
	WML		0.06
	CAC		0.04
	CMB	Global Cognition	0.20
	WML		0.05
	CAC		0.05
19q13-ADPRS	CMB	Memory	0.14
	WML		0.06
	CAC		0.03
	CMB	Global Cognition	0.20
	WML		0.05
	CAC		0.05
non-19q13-ADPRS	CMB	Memory	0.14
	WML		0.06
	CAC		0.04
	CMB	Global Cognition	0.21
	WML		0.05
	CAC		0.05

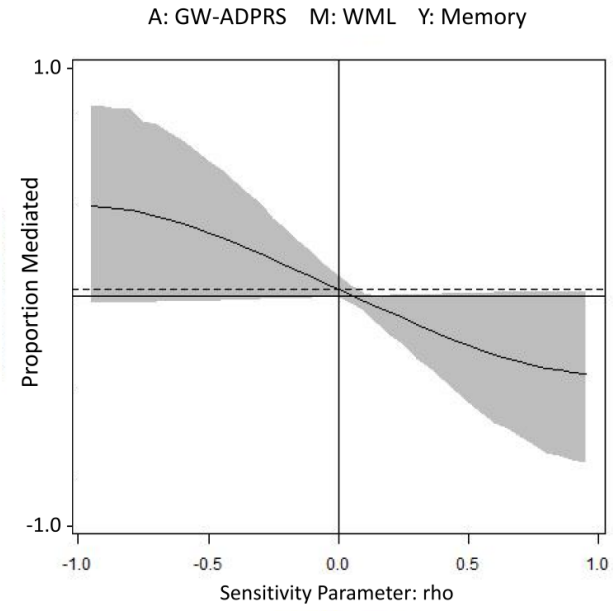
SUPPLEMENTARY FIGURES

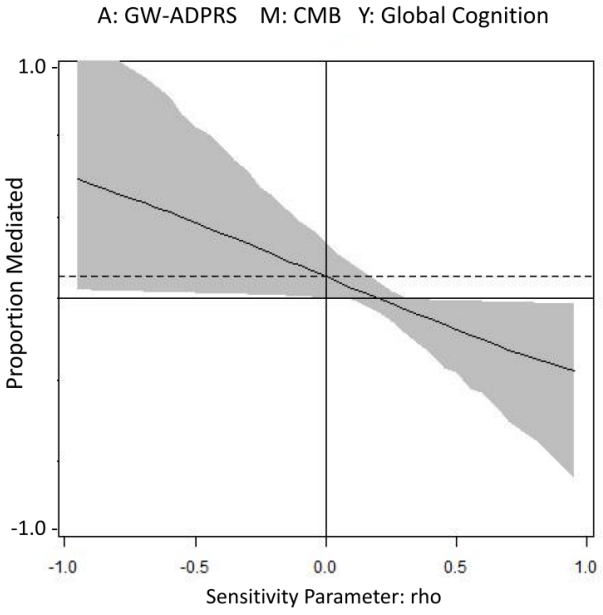
Figure S2.1: Sensitivity analyses of unmeasured confounding

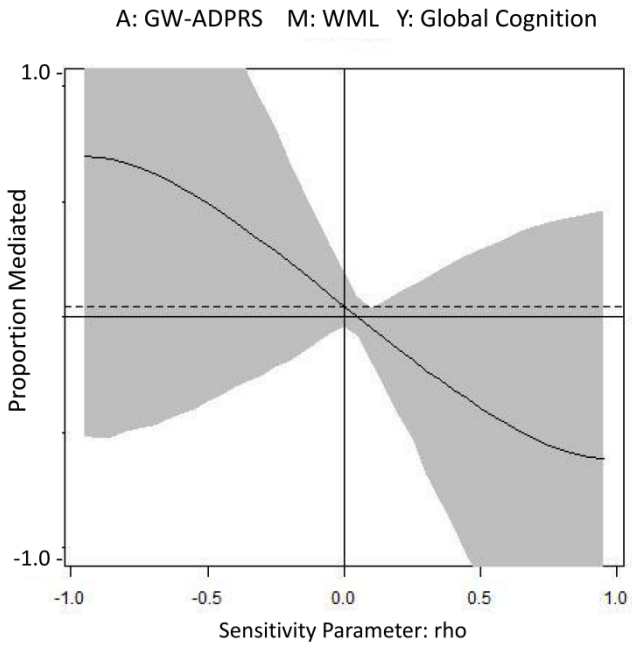
A: exposure; GW-ADPRS, 19q13-ADPRS, or non-19q13-ADPRS
M: mediator; CMB, WML, or CAC
Y: outcome; memory or global cognition score
The y-axis is the proportion mediated
The x-axis, denoted by rho, is the size of the correlation between the residuals in the equation predicting M and the equation predicting Y. We assume there is an unmeasured confounder variable that introduces this correlation between the residuals. The larger the absolute value of rho, the stronger the confounding.
The solid curve shows the estimated proportion mediated for different values of the correlation between the residuals in equations. The shaded part of the plot represents the 95% intervals surrounding the mediated effect. The x-intercept represents the value of rho at which proportion mediated equals to 0.
The horizontal broken line denotes the proportion mediated without considering unmeasured confounding. When rho is equal to zero, the reported proportion mediated is the same as that we estimated in the mediation analysis without considering unmeasured confounding. For other values of rho, the proportion mediated is calculated under different levels of unobserved confounding.
Models for the effects of GW-PRS adjusted for age, sex, smoking status, midlife physical activity, and diet quality.
Models for the effects of 19q13-PRS adjusted for age, sex, smoking status, midlife physical activity, diet quality, and non-19q13-PRS.
Models for the effects of non-19q13-PRS adjusted for age, sex, smoking status, midlife physical activity, diet quality, and 19q13-PRS.

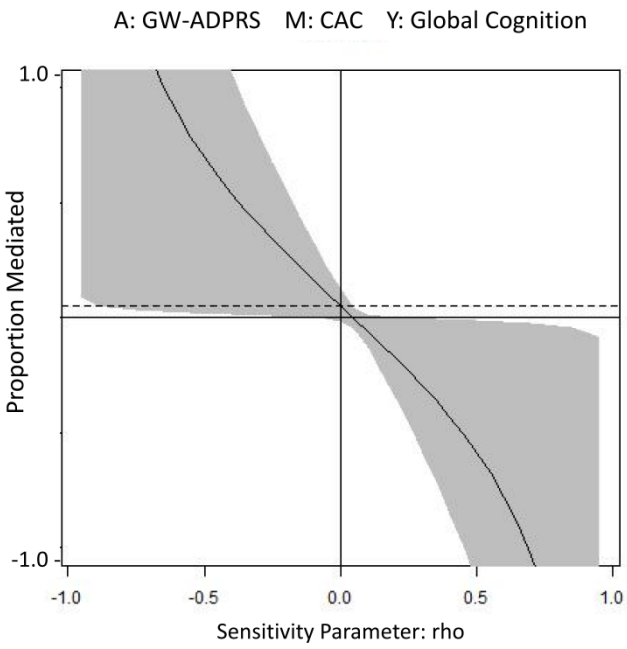

A: GW-ADPRS M: CMB Y: Memory
1.0
-1.0
Proportion Mediated
-1.0
-0.5
0.0
0.5
1.0
Sensitivity Parameter: rho

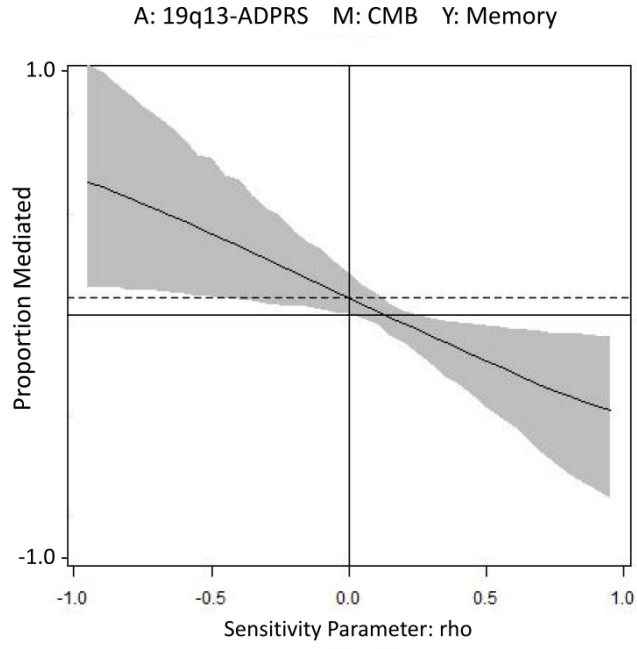

A: non-19q13-ADPRS M: CAC Y: Memory
1.0
-1.0
Proportion Mediated
-1.0
-0.5
0.0
0.5
1.0
Sensitivity Parameter: rho

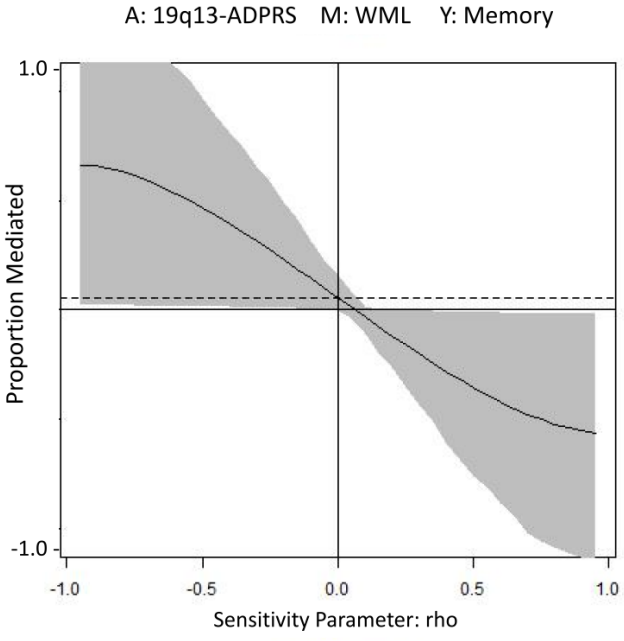

A: GW-ADPRS M: WML Y: Memory
1.0
-1.0
Proportion Mediated
-1.0
-0.5
0.0
0.5
1.0
Sensitivity Parameter: rho


A: GW-ADPRS M: CMB Y: Global Cognition
1.0
-1.0
Proportion Mediated
-1.0
-0.5
0.0
0.5
1.0
Sensitivity Parameter: rho

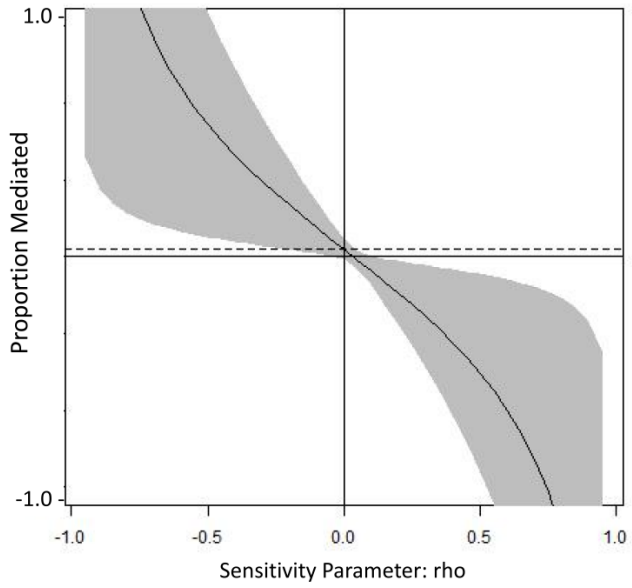

A: GW-ADPRS M: WML Y: Global Cognition
Proportion Mediated
1.0
-1.0
-1.0
-0.5
0.0
0.5
1.0
Sensitivity Parameter: rho


A: GW-ADPRS M: CAC Y: Global Cognition
Proportion Mediated
1.0
-1.0
-1.0
-0.5
0.0
0.5
1.0
Sensitivity Parameter: rho

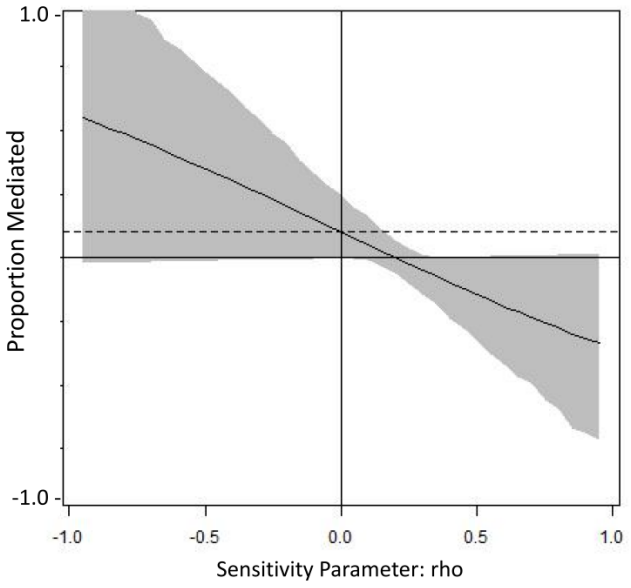

A: 19q13-ADPRS M: CMB Y: Memory
Proportion Mediated
1.0
-1.0
-1.0
-0.5
0.0
0.5
1.0
Sensitivity Parameter: rho


A: 19q13-ADPRS M: WML Y: Memory
Proportion Mediated
1.0
-1.0
-1.0
-0.5
0.0
0.5
1.0
Sensitivity Parameter: rho

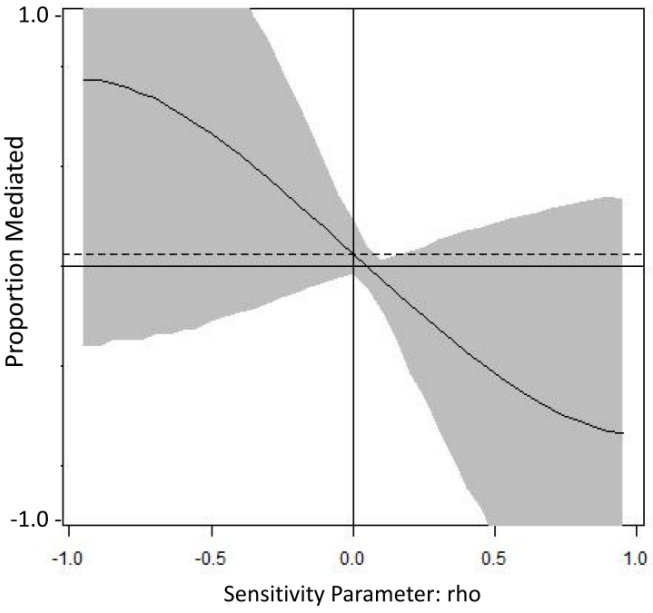
A: 19q13-ADPRS M: CAC Y: Memory



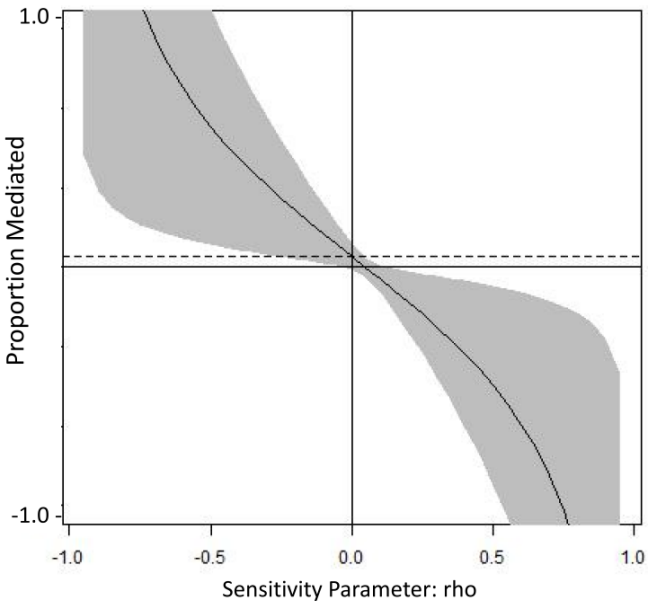
A: 19q13-ADPRS M: CMB Y: Global Cognition

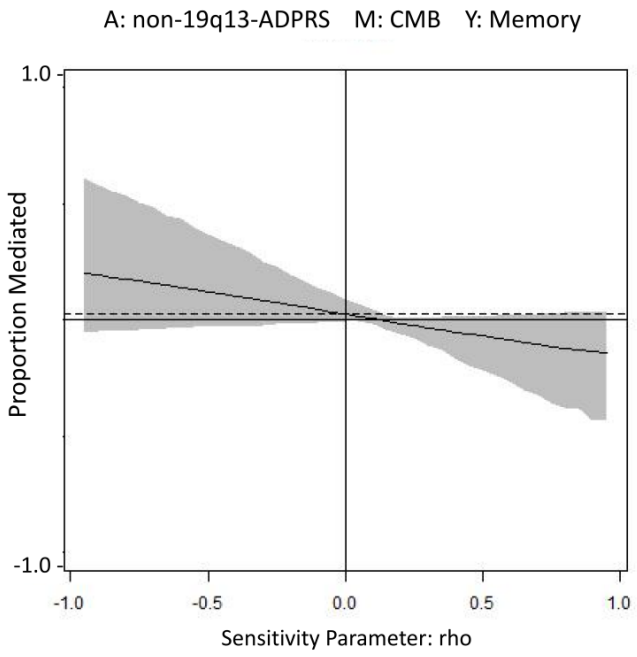


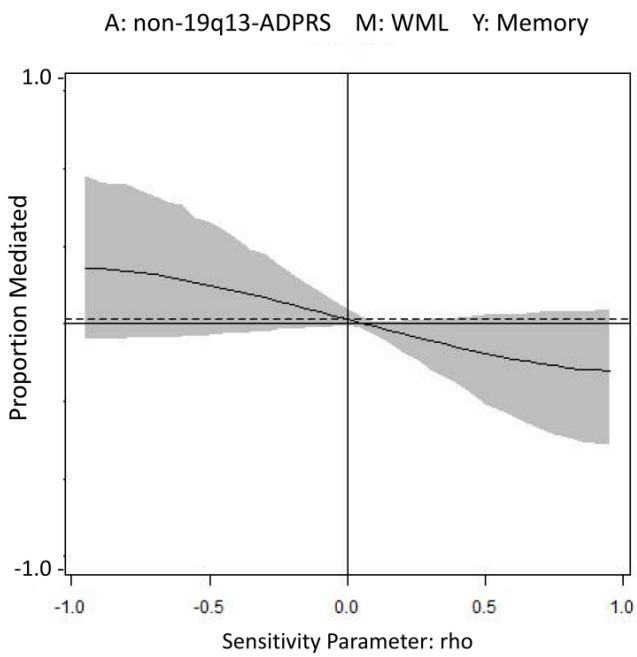
A: 19q13-ADPRS M: WML Y: Global Cognition

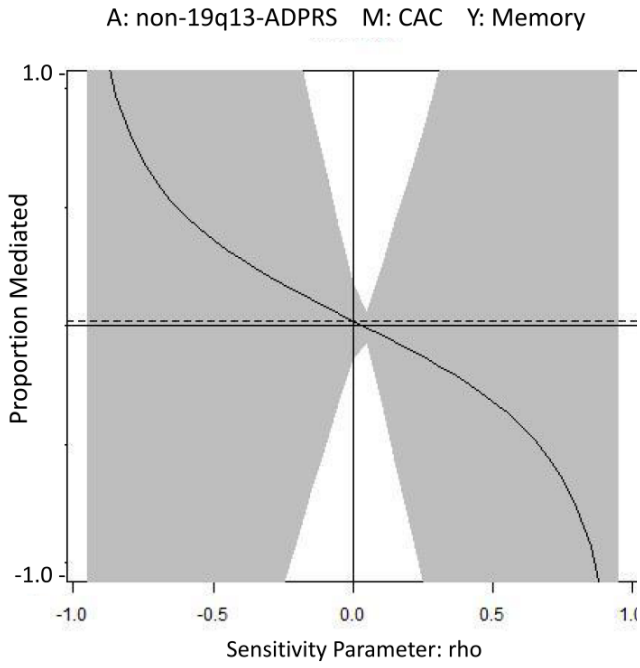


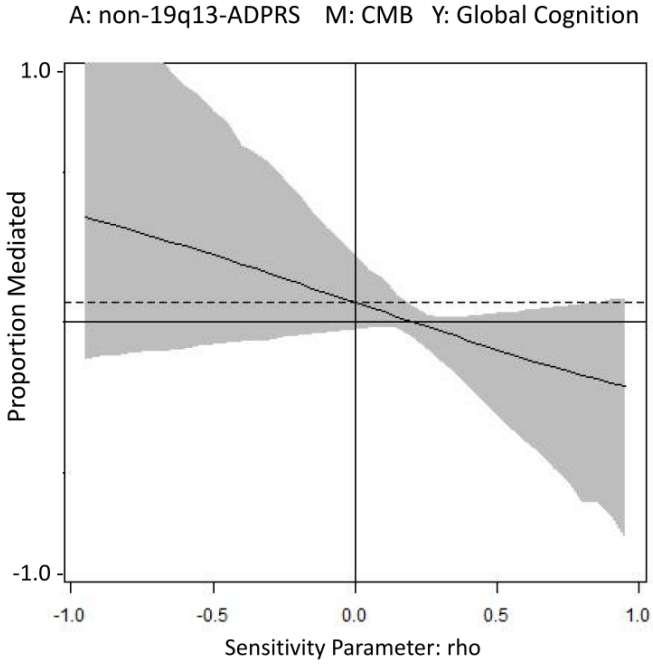
A: 19q13-ADPRS M: CAC Y: Global Cognition



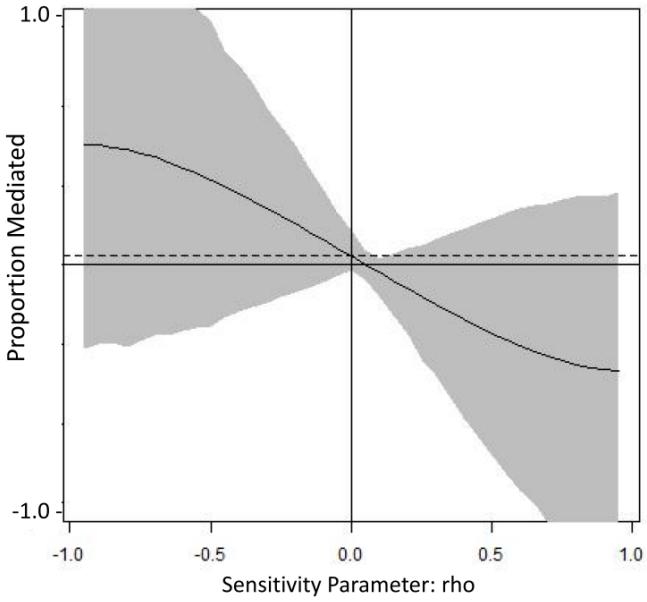

A: non-19q13-ADPRS M: CMB Y: Memory
1.0
-1.0
Proportion Mediated
-1.0
-0.5
0.0
0.5
1.0
Sensitivity Parameter: rho


A: non-19q13-ADPRS M: WML Y: Memory
1.0
-1.0
Proportion Mediated
-1.0
-0.5
0.0
0.5
1.0
Sensitivity Parameter: rho

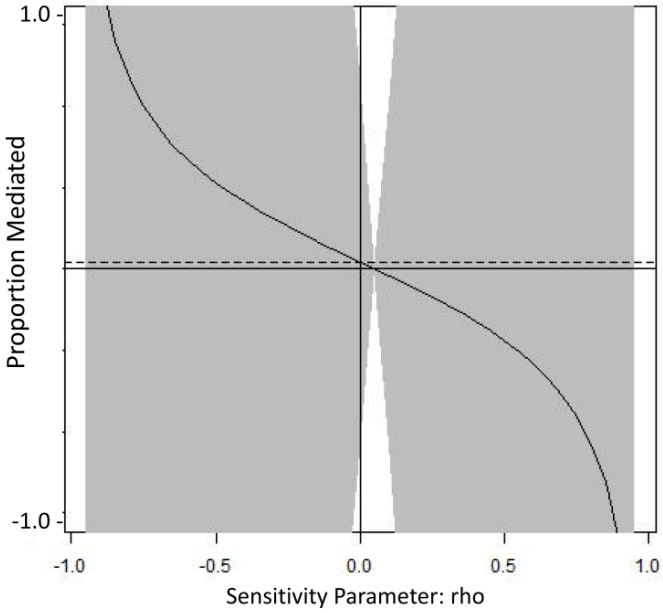

A: non-19q13-ADPRS M: CAC Y: Memory
1.0
-1.0
Proportion Mediated
-1.0
-0.5
0.0
0.5
Sensitivity Parameter: rho


A: non-19q13-ADPRS M: CMB Y: Global Cognition
1.0
-1.0
Proportion Mediated
-1.0
-0.5
0.0
0.5
1.0
Sensitivity Parameter: rho

A: non-19q13-ADPRS M: WML Y: Global Cognition



A: non-19q13-ADPRS M: CAC Y: Global Cognition



REFERENCES

Imai, K., Keele, L., Tingley, D., 2010. A general approach to causal mediation analysis. Psychol Methods 15, 309-334.

Imai, K.K., L.; Yamamoto, T., 2010. Identification, inference and sensitivity analysis for causal mediation effects. Statistical Science 25, 51-71.

Vanderweele, T.J., 2015. Explanation in causal inference : methods for mediation and interaction. . Oxford University Press, New York.

Chapter 3

Causal relationships between genetics, psychological traits and coronary heart disease

Yen-Feng Lin, Deborah Blacker, Alkes Price, Rebecca Betensky, and Jordan W. Smoller

ABSTRACT

OBJECTIVE:

Optimism/pessimism and cynical hostility have been linked to the risk of coronary heart disease (CHD). In this study, we examined the genetic overlap between these psychological traits and CHD, and explored the extent to these traits mediated genetic influences on CHD risk.

METHODS:

For 6336 African American (AA), 13735 European American (EA), and 2935 Hispanic American (HA) relatively healthy women with genotype data available in the Women's Health Initiative (WHI) Study, we generated genome-wide polygenic risk scores for CHD (CHD-PRS), and examined the association of CHD-PRS with optimism/pessimism and cynical hostility, respectively. Where associations were observed, we performed mediation analyses to determine whether the association of CHD-PRS with CHD risk is mediated by these traits.

RESULTS:

In the EA sample, CHD-PRS was significantly associated with optimism (t=-3.43, P=6.09E-04, R^2=0.0009) and liability to both all CHD (t=5.87, P=4.51E-09, R^2=0.0025) and acute CHD (t=4.38, P=1.20E-05, R^2=0.0014), but was not associated with cynical hostility. In the AA and HA samples, CHD-PRS was associated with liability to CHD but neither optimism nor cynical hostility. In the EA sample, the total effect of CHD-PRS on liability to CHD was significantly mediated by optimism (proportion mediated=1.4% for all CHD and proportion mediated=1.7% for acute CHD).

CONCLUSION:

Despite several limitations, our findings suggest a genetic overlap between optimism and CHD in older women of European ancestry. The polygenic genetic effect on CHD is modestly though significantly mediated by optimism.

INTRODUCTION

Coronary heart disease (CHD) is the most common type of heart disease, responsible for more than 370,000 deaths annually in the US (NCHS., 2015). (Nikpay et al., 2015)Psychological factors have been linked to the risk of developing CHD (Albus, 2010). In particular, dimensions of negative affectivity, including anger, hostility, and pessimism have been implicated in CHD risk. For example, several prospective cohort studies have found that cynicism, the cognitive component of hostility, is associated with increased risk of CHD (Barefoot, Larsen, von der Lieth, & Schroll, 1995; Barefoot et al., 1991; Izawa et al., 2011; Wong, Na, Regan, & Whooley, 2013). A meta-analysis of prospective observational studies found that hostility was associated with an increased CHD risk in healthy populations and with poor prognosis in individuals with CHD (Chida & Steptoe, 2009a). In addition, epidemiological evidence supports an association of dispositional optimism or pessimism with CHD (Hansen et al., 2010; Kubzansky, Sparrow, Vokonas, & Kawachi, 2001; Pankalainen, Kerola, & Hintikka, 2015; Tindle et al., 2009).

There are several possible ways in which cynical hostility and optimism/pessimism could directly contribute to the development of CHD. Cynical hostility may increase the risk of CHD via dysregulated autonomic nervous system (Chida & Hamer, 2008; Thomas, Nelesen, & Dimsdale, 2004; Vella & Friedman, 2007), high cortisol levels (Steptoe, Cropley, Griffith, & Kirschbaum, 2000), and elevated biomarkers of inflammation and coagulation (Markovitz, 1998; Stewart, Janicki-Deverts, Muldoon, & Kamarck, 2008). It is also possible that cynical hostility alters susceptibility to CHD via behavioral pathways related to unhealthy lifestyle, smoking, sleep problems, or poor adherence to treatment (Scherwitz et al., 1992; Shin et al., 2005; Siegler,

Peterson, Barefoot, & Williams, 1992). Similarly, pessimism has been shown to be associated with impaired stress coping (Scheier, Weintraub, & Carver, 1986), autonomic dysfunction (Raikkonen, Matthews, Flory, Owens, & Gump, 1999; Sharot, Riccardi, Raio, & Phelps, 2007), higher cortisol awakening response (Chida & Steptoe, 2009b), elevated inflammatory markers (O'Donovan et al., 2009; Roy et al., 2010), and thrombogenesis (Roy et al., 2010). Pessimists may also tend to have more unhealthy behaviors, such as smoking and less physical activity (Steptoe, Wright, Kunz-Ebrecht, & Iliffe, 2006; Tindle et al., 2009), and poor medical adherence (Tinker et al., 2007). Although these physiological and behavioral pathways may explain a possible causal relationship between psychological traits and CHD, the observed associations of CHD with optimism/pessimism and cynical hostility can be confounded by common underlying factors.

Genetic studies may give useful information to explain the relationships between psychological traits and CHD. It is now well-established that genetic variation contributes to the risk of CHD. Genome-wide association studies (GWAS) have identified more than 56 common genetic variants that contribute to the heritability of CHD (Nikpay et al., 2015). With the development of behavioral genetics, there is growing interest in the genetics of psychological and personality traits. Recently, GWAS have been used to identify genetic variants of hostility and proneness to anger (Merjonen et al., 2011; Mick et al., 2014). Although no GWAS has been reported for optimism/pessimism, a twin/adoption study estimated the heritability of optimism and pessimism at 25% (Plomin, 1992). In addition, a recent GWAS suggested that extraversion of the Big Five Model of Personality (McCrae & Costa, 1987), which has been shown to negatively

correlate with pessimism (Mahasneh, 2013), is a highly polygenic trait (van den Berg et al., 2016).

Methods have been developed recently to investigate overlap between polygenic traits using GWAS data (Andreassen et al., 2013; Bulik-Sullivan et al., 2015; International Schizophrenia et al., 2009), and their applications have provided insights into the overlapping genetics of numerous traits and disorders. However, to date there has been no study reporting a shared genetic basis between CHD and psychological traits. Given the knowledge that both cynical hostility and optimism/pessimism are associated with CHD, and all of them are heritable complex traits, a logical next step is to examine the genetic correlation between CHD and each of the psychological traits. Moreover, it is worthwhile to further explore the causal relationship between genetic factors, psychological traits, and CHD. Elucidation of the causal mechanism can provide insight into the potential effectiveness of limiting genetic effects on CHD by intervening on cynical hostility or pessimism.

In the present study, we hypothesized that there may be polygenic overlap between psychological traits and the development of CHD, and some of the genetic effects on CHD may be mediated by psychological traits. We analyzed data from the Women's Health Initiative (WHI) GWAS genotyped samples to test the above hypotheses. With the availability of large-scale data sets that combine genome-wide data with psychological measures and CHD outcomes, we examined the SNP-based genetic overlap between psychological traits (cynical hostility and optimism/pessimism) and CHD. We also performed a causal mediation analysis

and quantify how much of the polygenic effect on CHD is mediated by optimism/pessimism or cynical hostility.

METHODS

Study Population

The Women's Health Initiative Observational Study (WHI-OS) and Clinical Trials (WHI-CT) were carried out at 40 US clinical centers between 1993 and 1998 ("Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group," 1998; Langer et al., 2003). Participants were postmenopausal women aged 50 to 79 years at baseline, free from serious cardiac, pulmonary, renal, hepatic, and mental illness, with at least 3 years' life expectancy. The WHI-CTs performed randomized controlled trial evaluation of three distinct interventions: a low-fat eating pattern, menopausal hormone therapy (HT), and calcium plus vitamin D supplementation. The WHI-OS was designed to provide information about disease risk factors, including cancer, cardiovascular disease, and fractures. The combined studies enrolled 161,808 participants (93,676 in the WHI-OS and 68,132 in the WHI-CTs).

In the present analyses, we used phenotype and genotype data from the six independent GWAS sub-studies within the WHI study (Supplementary Table S2.1): (1) the SNP Health Association Resource cohort (SHARe) (n=7,470 African Americans [AA] and 3,348 Hispanic Americans [HA]), (2) the Genome-Wide Association Studies of Treatment Response in Randomized Clinical Trials (GARNET) (n = 3,727 European Americans [EA]), (3) the WHI Memory Study (WHIMS+) (n=5,687 EA), (4) the Hip Fracture GWAS (HIPFx) (n=2,841 EA), (5) the Modification of PM-Mediated Arrhythmogenesis in Populations (MOPMAP) (n=2,840 EA), and (6) the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO) (n=2,083 EA).

Genotyping and Quality Control

All sub-studies performed genome-wide genotyping with either Affymetrix (Santa Clara, CA, USA) or Illumina (San Diego, CA, USA) arrays, and excluded SNPs and samples with similar quality control filters (Supplementary Table S2.1). DNA samples and SNPs with call rates less than 90–98%, depending on the sub-study were excluded. Within each sub-study, SNPs with minor allele frequencies (MAF) less than 1% or that failed Hardy-Weinberg equilibrium (P <1E-04 or 1E-06) were excluded (Supplementary Table S2.2). Sex mismatches, duplicate samples, one of the first- and second- degree relatives, and subjects whose genetic ethnicity was inconsistent with their self-reported ethnicity were excluded. Genotypes were imputed using 1000 Genomes Phase 1 reference panels, with the MaCH/ minimac program (Howie, Fuchsberger, Stephens, Marchini, & Abecasis, 2012) (Supplementary Table S2.2). We filtered imputed SNPs based on imputation quality > 0.1.

For the present analyses, we excluded subjects with a history of stroke, myocardial infarction (MI), angina, percutaneous transluminal coronary angioplasty (PTCA), coronary artery bypass graft (CABG), or congestive heart failure (CHF) at baseline. The final study samples consisted of 6336 AA, 13735 EA, and 2935 HA subjects. All the analyses were performed separately for AA, EA and HA WHI subjects.

Measurement of Optimism and Hostility

At the baseline examination, psychological traits (optimism and cynical hostility) were measured by questionnaires for all participants. Optimism was assessed by 6 items of the Life

Orientation Test–Revised (Scheier, Carver, & Bridges, 1994). Item ratings were summed to yield a total score that ranges from 6 to 30 (higher scores indicate greater optimism, and lower scores indicate greater pessimism). Cynical hostility was measured by the cynicism subscale of the Cook-Medley Questionnaire (Cook, 1954), containing 13 true/false questions. The cynicism score ranges from 0 to 13, with a higher score indicating greater cynical hostility. The correlation between optimism and hostility was r= − 0.27 (P<0.0001) for the AA subsample, r= − 0.30 (P<0.0001) for the EA subsample, and r= − 0.29 (P<0.0001) for the HA subsample.

Measurement of CHD

Outcomes were adjudicated through 2010 according to a previously described protocol (Curb et al., 2003). Incidence of all CHD events (defined as MI, angina, coronary revascularization, or CHD death) and acute CHD events (defined as MI or CHD death) were adjudicated locally and centrally from medical records including hospital discharge summaries, ICD-9 codes, diagnostic, laboratory, surgical, and pathology reports by trained physicians blinded to randomized intervention and exposure status.

Since age is one of the most important risk factors for CHD and may also modify genetic risk, we used an informed conditioning approach (Zaitlen et al., 2012) to account for the influence of age on the genetic risk for CHD. The informed conditioning method uses a liability threshold model with parameters informed by published epidemiological statistics, accounting for clinical covariates, disease prevalence, and non-random ascertainment of cases in case-control association studies. In brief, we computed a conditional liability to all CHD for each individual,

based on one's case-control status (i.e., whether having a CHD event during the follow-up), conditioning on age (the age at first CHD event for cases, and the age at end of follow-up for non-cases) and considering previously reported age-stratified prevalence of CHD in the U.S. (Mozaffarian et al., 2015). We also used the same method to compute a conditional liability to acute CHD for each subject. The age-conditional liability to all CHD and acute CHD were the main outcomes of the present study.

Measurement of Other Baseline Characteristics

Other baseline covariates used in the analyses included: age (years, continuous), education (less than a high school education, greater than high school but less than college, any college, or more than college), living alone (binary, yes or no), annual household income (less than $20000, $20000-$49999, $50000-$99999, $100000 or more), religious service attendance (not at all, 1 to 3 times per month, or ≥ 1 time per week), diabetes (binary, yes or no), hypertension (binary, yes or no), hypercholesterolemia (binary, yes or no), smoking (never, ever, or current), alcohol consumption (never, ever, or current), physical activity (< 2.5, 2.5 to < 18.25, or ≥ 18.25 metabolic equivalents), large waist circumference (binary, ≥ 88 or <88 cm), body mass index (continuous), and OS/CT status (binary, observational study or clinical trial).

Statistical Analysis

Phenotypic associations of psychological traits with CHD

We used multiple linear regression to examine the associations of each psychological trait with the age-conditional liability to all CHD and acute CHD, adjusting for age, education, living alone,

annual household income, religious service attendance, diabetes, hypertension, hypercholesterolemia, smoking, alcohol consumption, physical activity, waist circumference, body mass index, OS/CT status. We performed the analyses separately in each ethnic subsample.

PRS for CHD in the European Ancestry sample

We used the summary statistics from the Coronary ARtery DIsease Genome wide Replication and Meta-analysis consortium (CARDIoGRAM) GWAS (22,233 CHD cases and 64,762 controls of European ancestry) (Schunkert et al., 2011) as the training dataset to derive genome-wide polygenic risk scores for CHD (CHD-PRS) in our WHI-EA sample. To include only independent association signals from the CARDIoGRAM GWAS, we applied an LD clumping procedure to the training data in which we retained the SNP with smallest P-value in each 250kb window and removed all those in LD (r^2> 0.2) with this SNP. We used three different association P-value thresholds (PTs), 0.0001, 0.001, and 0.01, to select index SNPs from the clumped independent SNPs for generating the PRSs. For each individual, and each P-value threshold (PT), we calculated a CHD-PRS by summing the risk allele counts of the index SNPs, weighted by the log of their association odds ratios estimated from the CARDIoGRAM GWAS results.

PRS for CHD in the African Ancestry and Hispanic Ancestry samples

The ideal training dataset for PRS construction in the AA and HA samples would be large-scale GWAS from the same populations, but such GWAS were not available. In order to improve the PRS prediction between ethnically diverse samples, we used a newly developed method

MultiPRS (Marquez Luna et al., 2016), combining the training GWAS data from the European CARDIoGRAM and cross-validations within each of our WHI AA and HA samples. First, we used CARDIoGRAM GWAS results as the training dataset to calculate PRS with European training data (EUR-PRS) with different p-value thresholds (0.0001, 0.001, and 0.01) in each of AA and HA subsamples. Second, we used a five-fold cross-validation procedure to calculate PRS with AA training data (AA-PRS) and PRS with HA training data (HA-PRS) with a p-value threshold of 0.0001, 0.001, or 0.01 in each ethnic subsample. The procedures for construction of EUR-PRS and each TARGET-PRS (i.e., AA-PRS or HA-PRS) were similar to those for generating CHD-PRS in the EA subsample, but with different training and target datasets. Third, we built prediction models for the CHD outcomes (i.e., the conditional liability to all CHD and acute CHD) using the best linear combination of EUR-PRS, TARGET-PRS, and top 10 principal components (PCs) of ancestry of the target AA or HA subsamples. We performed a 2-dimensional grid search to optimize the linear combination via an in-sample fit on validation samples, with different P-value thresholds for EA-PRS and TARGET-PRS. We examined the prediction accuracy (adjusted R^2) of the 9 (3x3) combinations for each ethnic subsample, and used the fitted beta coefficients as the weights to generate the final CHD-PRS for each AA and HA participant.

CHD-PRS associations with CHD and each psychological trait in target samples

As described above, we generated CHD-PRS with different p-value thresholds for each of the AA, EA, and HA samples. We then examined their association with the conditional liability to CHD within each ethnic sample. We also examined associations between the CHD-PRS and both either optimism and cynicism, respectively, within each ethnic sample. For the EA

subsample, associations were tested using linear regression models with the top 10 PCs of ancestry as covariates. For the AA and HA subsamples, we did not adjust for PCs because the information of PCs was already incorporated into the PRS. The Wald test P-value for each association test was reported, and squared semi-partial correlations (for EA) and adjusted R squared (for AA and HA) were calculated to estimate the proportion of variance explained by the PRSs.

Causal Mediation Analyses

We then explored the causal relationship between CHD-PRS, psychological traits, and liability to all CHD or acute CHD. We performed causal mediation analyses to examine how much of the effect of a CHD-PRS on the age-conditional liability to CHD was mediated by a psychological trait shown to be associated with CHD-PRS. We used a method based on the counterfactual framework for causal inference (Pearl, 2001; Robins & Greenland, 1992), which is an extension of traditional regression-based mediation approaches (Baron & Kenny, 1986), allowing exposure-mediator interactions (T. J. VanderWeele, 2016).

We estimated the direct and indirect (mediated) effects with CHD-PRS as the exposure, a psychological trait (if observed to be associated with CHD-PRS) as the mediator, and the age-conditional liability to all CHD or acute CHD as the outcome. To maximize statistical power, mediation analyses were conducted using the PRS P-value threshold that showed the strongest association with the CHD outcome. We adjusted for the top 10 PCs of ancestry and potential mediator-outcome confounders, including age, living alone, annual household income, religious

service attendance, alcohol consumption, physical activity, and OS/CT status. A counterfactual outcome variable denotes the outcome that would have been observed had an exposure been set to a particular value. In order to broadly understand the impact of the CHD-PRS as an exposure, we selected its 25^{th} and 75^{th} percentile to compare when we estimated the direct and the indirect effects. As sensitivity analyses for our selection of exposure levels for comparison, we also repeated our mediation analyses comparing the effects of the 90th percentile to the 10th percentile of the CHD-PRS. As an index of mediated effects, we calculated the proportion mediated (PM) by dividing the estimated indirect effect by the estimated total effect.

All the mediation analyses were performed using the PARAMED module in STATA (Emsley & Liu, 2013; Valeri & Vanderweele, 2013). We used bootstrap procedures with 200 replications to compute a 95% bias-corrected bootstrap confidence interval (95% BCCI) for the direct and indirect effects. Finally, we conducted sensitivity analyses to evaluate the robustness of the above mediation analyses to unmeasured confounding (see Supplementary Methods).

RESULTS

Sample Characteristics

Table 3.1 presents descriptive statistics of baseline and follow-up characteristics for the WHI-AA, EA, and HA samples. Among the 6,336 AA participants, 507 (8.0%) had a CHD event (including 241 [3.8%] with acute CHD) during the follow-up period (average = 5409 days, range= 1811 to 7154 days). Among the 13,735 EA participants, 1,664 (12.1%) had a CHD event (including 981 [7.1%] with acute CHD) during the follow-up follow-up period (average = 5404 days, range=254 to 7125 days). Among the 2,935 HA participants, 192 (6.5%) had a CHD event (including 80 [2.7%] with acute CHD) during the follow-up follow-up period (average = 5267 days, range= 1802 to 7052 days).

Phenotypic Associations

Table 3.2 presents phenotypic associations between each psychological trait and the conditional liability to CHD. After adjusting for potential confounders, optimism was significantly associated with the liability to both all CHD and acute CHD, while cynicism was significantly associated with the liability to acute CHD in the EA subsample. No association between psychological traits and CHD outcomes was observed in the AA and HA subsamples.

Associations of CHD-PRS with the liability to CHD and psychological traits

We performed PRS association analyses to examine the association between CHD-PRS and each psychological triat, and the association between CHD-PRS and the conditional liability of all CHD or acute CHD.

EA subsample

The results of PRS association analyses for EA participants are presented in Table 3.3. The CHD-PRS with a p-value threshold of 0.0001 ($CHD\text{-}PRS_{PT = 0.0001}$) was significantly associated with optimism (t=-3.43, P=6.09E-04, R^2=0.0009) and the liability to both all CHD (t=5.87, P=4.51E-09, R^2=0.0025) and acute CHD (t=4.38, P=1.20E-05, R^2=0.0014). None of the CHD-PRS was associated with cynical hostility.

AA and HA subsamples

The results of PRS association analyses for AA and HA participants are presented in Supplementary Table S3.3 and Table S3.4, respectively. For the AA subsample, the CHD-PRS with a p-value threshold of 0.0001 for EUR-PRS and a p-value threshold of 0.01 for AA-PRS was significantly associated with the liability to both all CHD (t=3.31, P=7.20E-04, R^2=0.0011) and acute CHD (t=2.98, P=1.53E-03, R^2=0.0009). For the HA subsample, the CHD-PRS with a p-value threshold of 0.0001 for EUR-PRS and a p-value threshold of 0.001 for HA-PRS was nominally associated with the liability to both all CHD (t=2.47, P=0.01, R^2=0.0007) and acute CHD (t=2.01 P=0.04, R^2=0.0004). However, in the WHI-AA and HA samples, none of the CHD-PRS was associated with optimism or cynical hostility.

Mediation Analyses

Our PRS association analyses suggested possible SNP-based genetic overlap between optimism and CHD in EA individuals. To examine whether the effect of CHD-PRS on the liability of CHD

was mediated through optimism, we performed regression-based causal mediation analyses to decompose the total effect of PRS on each of the CHD outcomes (i.e., the age-conditional liability of all CHD and acute CHD) into direct and indirect (mediated) effects, adjusting for potential confounders. The CHD-PRS with a p-value threshold of 0.0001 ($CHD\text{-}PRS_{PT = 0.0001}$) was the most associated with the CHD outcomes, and was selected as the exposure in the mediation analyses. Optimism score, which was also observed to be associated with $CHD\text{-}PRS_{PT = 0.0001}$ was tested as a potential mediator.

The estimated direct and indirect effects on the liability to CHD, comparing the 75th percentile versus the 25th percentile and the 90th percentile versus the 10th percentile of the CHD-PRS, are shown in Table 3.4. The total effect of CHD-PRS on the liability of CHD was significantly mediated by optimism (PM=1.4% for all CHD and PM=1.7% for acute CHD). Different exposure levels appeared to yield similar PMs. When an interaction between the CHD-PRS and optimism was included in each mediation model, there was very little change in the estimated direct and indirect effects, so these were not included in the mediation models, as suggested by Vanderweele (T.J. Vanderweele, 2015).

Sensitivity analyses to assess the effect of unmeasured confounding suggested that in the presence of an unmeasured confounder associated with higher risk of CHD and greater optimism or an unmeasured confounder associated with lower risk of CHD and less optimism, our estimated PMs would underestimate the true mediation effects of optimism. For example, if there was unmeasured confounding with correlations of 0.1 with optimism score and of 0.1

with liability to all CHD, the PM would increase from 1.4% to 1.9%. However, if the associations of the unmeasured confounder with optimism and CHD risk were in different directions, our estimated PMs would overestimate the true mediation. For example, if there was unmeasured confounding with correlations of -0.1 with optimism score and of 0.1 with liability to all CHD, the PM would reduce from 1.4% to 1.0% (see Supplementary Tables S3.5 and S3.6).

DISCUSSION

Summary of Main Findings

Among relatively healthy post-menopausal women of European ancestry in the WHI (WHI-EA participants), we found that greater optimism was significantly associated with a lower risk of later acute CHD events and all CHD end points. We also observed that cynical hostility was significantly associated with acute but not chronic CHD events in WHI-EA participants. We also observed a significant association between CHD-PRS and optimism in the WHI-EA sample (P =6.09E-04), suggesting some degree of genetic overlap between CHD and optimism. Mediation analyses in the WHI-EA subsample showed that a small but significant portion of the effects of CHD-associated alleles on both acute and chronic CHD were mediated by optimism. Neither phenotypic nor genetic associations between psychological traits and CHD was observed in WHI-AA and HA subsamples.

Phenotypic Associations between Psychological Traits and CHD

Numerous studies have linked CHD to optimism/pessimism (Hansen et al., 2010; Kubzansky et al., 2001; Pankalainen et al., 2015) and cynical hostility (Barefoot et al., 1995; Barefoot et al., 1991; Izawa et al., 2011; Wong et al., 2013), including in the WHI itself (Tindle et al., 2009). Our phenotypic analyses (Table 3.2) replicated previously reported associations of optimism with both acute and chronic CHD and an association of cynical hostility with acute CHD These associations were observed in European American but not African American women, consistent with the results of a previous study, using all WHI EA and AA subcohorts with or without GWAS data available (Tindle et al., 2009). We also examined the relationship between psychological

traits and CHD in women of Hispanic ancestry, and did not detect a significant association. Our findings suggest that the impact of psychological traits on CHD risk may differ between ethnic populations. It is possible that the non-significant findings in AA and HA subsamples are due to the smaller sample sizes, but this is less likely because the point estimates for the effect (i.e., the beta coefficients) are small or even opposite in direction.

Genetic Overlap between Optimism and CHD

As noted above, consistent with prior studies, we observed significant associations between psychological traits of optimism and cynical hostility with CHD. However, whether these phenotypic links represent underlying common biological etiology had not previously been clarified. Using a polygenic approach to GWAS data, we were able to further examine if shared genetic factors contribute to the observed phenotypic associations between psychological traits and CHD. We found an association between optimism and CHD-PRS, with the strongest signal at a P-value threshold of 0.0001, in European American women, suggesting that the relationship between CHD and optimism reflects shared genetic influences. Nevertheless, in the same WHI-EA sample, there was no genetic correlation between cynical hostility and CHD, implying the observed association between cynicism and CHD may be explained by non-genetic common factors.

Among AA and HA participants, a trans-ethnic CHD-PRS was not associated with optimism or cynical hostility in the WHI-AA and HA subsamples. The small effect sizes in these analyses suggest that the smaller sample sizes are unlikely to account for the non-significant findings.

Optimism Partially Mediates the Polygenic Effect on CHD

Having established that polygenic risk of CHD is associated with both optimism and CHD in European American women, we performed causal mediation analyses to assess whether this cross-trait genetic relationship is due to biological pleiotropy (a causal variant directly affecting both traits) or mediated pleiotropy (a causal variant affecting one trait, which in turn affects the other trait) (Solovieff, Cotsapas, Lee, Purcell, & Smoller, 2013). We found that a small but measurable portion of the CHD-PRS effect on CHD risk was mediated by optimism in non-Hispanic white women. Our findings indicate that CHD-associated SNPs affect the risk of CHD partially through a pathway involving optimism/ pessimism, suggesting that an intervention that influences the degree of optimism or pessimism may be effective for the prevention of CHD. A recent meta-analysis on 29 studies with a total of 3319 participants indicate that psychological interventions, such as the "best possible self" intervention, self-compassion training, cognitive-behavioral therapy, and positive-psychology methods, can significantly increase optimism (Malouff, 2017). Future research and prevention efforts on these psychological interventions may have the potential to develop strategies for lowering the risk of CHD for individuals with higher genetic loading.

Potential Causal Relationship from Psychological Traits to CHD

The established causal mediation relationship between CHD-PRS, optimism, and liability to CHD suggest some causal effect of optimism on CHD risk (Figure 3.1). In the analysis that identified phenotypic associations between optimism and CHD risk (Table 3.2), although we controlled for

a number of potential confounders, we cannot completely rule out the possibility of residual confounding. As shown in Figure 3.1, there are causal arrows from CHD-PRS to both optimism and CHD, indicating that CHD-PRS may also confound the causal relationship between optimism and CHD. In addition to the covariates included in the phenotypic analyses shown in Table 3.2, we could also adjust for the CHD-PRS to account for possible residual confounding by shared genetic factors. After additionally adjusting for CHD-PRS, optimism was still significantly associated with both the conditional liability to all CHD (beta=-0.0041, SE=0.0018, P=0.03) and acute CHD (beta=-0.0033, SE=0.0016, P=0.04).

With respect to the relationship between cynical hostility and CHD risk, we observed evidence of a phenotypic association but no genetic correlation. These findings do not rule out a causal relationship between the two phenotypes that could be mediated by other genetic or non-genetic factors. Further research could clarify this possibility. For example, if robust associations between genetic variants and cynicism are found in the future, these variants could serve as genetic instruments to test a causal relationship between cynicism and CHD in a mendelian randomization framework (Figure 3.2) (Smith & Ebrahim, 2003; Solovieff et al., 2013).

Limitations

The present study has several limitations. First, the PRS method tests only the effect of common genetic variants and the additive genetic model of inheritance. Without considering rare variants, gene-gene and gene-environment interactions, the genetic contribution captured by PRS may underestimate the true genetic effect on phenotypes. Second, the results of

mediation analyses might be biased due to violation of the unmeasured confounding assumption (T. J. VanderWeele, 2016). Our sensitivity analyses suggest that the evidence for mediation can be falsely detected with the existence of strong unmeasured confounders correlated with the mediator and the outcome in opposite directions. Finally, the WHI includes relatively healthy, post-menopausal women, thus our findings may not generalize to men and premenopausal women.

CONCLUSION

Using large-scale genomic resources, we observed genetic overlap between optimism and CHD risk among older women. In women of European ancestry, optimism appeared to partly mediate the relationship between CHD polygenic risk and CHD outcomes. Our results also indicate that cynical hostility, though phenotypically associated with acute CHD, may not share common genetic etiology with CHD, suggesting non-heritable shared environmental causes may play a more important role in the relationship between cynical hostility and CHD. Identifying the psychological traits associated with CHD risk in women will facilitate the identification of vulnerable women, and may also inform the development of effective prevention and intervention strategies to improve cardiovascular health among women.

REFERENCES

Albus, C. (2010). Psychological and social factors in coronary heart disease. *Ann Med, 42*(7), 487-494. doi:10.3109/07853890.2010.515605

Andreassen, O. A., Djurovic, S., Thompson, W. K., Schork, A. J., Kendler, K. S., O'Donovan, M. C., . . . Dale, A. M. (2013). Improved detection of common variants associated with schizophrenia by leveraging pleiotropy with cardiovascular-disease risk factors. *Am J Hum Genet, 92*(2), 197-209. doi:10.1016/j.ajhg.2013.01.001

Barefoot, J. C., Larsen, S., von der Lieth, L., & Schroll, M. (1995). Hostility, incidence of acute myocardial infarction, and mortality in a sample of older Danish men and women. *Am J Epidemiol, 142*(5), 477-484.

Barefoot, J. C., Peterson, B. L., Dahlstrom, W. G., Siegler, I. C., Anderson, N. B., & Williams, R. B., Jr. (1991). Hostility patterns and health implications: correlates of Cook-Medley Hostility Scale scores in a national survey. *Health Psychol, 10*(1), 18-24.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Pers Soc Psychol, 51*(6), 1173-1182.

Bulik-Sullivan, B., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., Loh, P. R., . . . Neale, B. M. (2015). An atlas of genetic correlations across human diseases and traits. *Nat Genet, 47*(11), 1236-1241. doi:10.1038/ng.3406

Chida, Y., & Hamer, M. (2008). Chronic psychosocial factors and acute physiological responses to laboratory-induced stress in healthy populations: a quantitative review of 30 years of investigations. *Psychol Bull, 134*(6), 829-885. doi:10.1037/a0013342

Chida, Y., & Steptoe, A. (2009a). The association of anger and hostility with future coronary heart disease: a meta-analytic review of prospective evidence. *J Am Coll Cardiol, 53*(11), 936-946. doi:10.1016/j.jacc.2008.11.044

Chida, Y., & Steptoe, A. (2009b). Cortisol awakening response and psychosocial factors: a systematic review and meta-analysis. *Biol Psychol, 80*(3), 265-278. doi:10.1016/j.biopsycho.2008.10.004

Cook, W. W. M., D. M. (1954). Proposed hostility and pharisaic-virtue scales for the MMPI. *Journal of Applied Psychology, 38*, 414-418.

Curb, J. D., McTiernan, A., Heckbert, S. R., Kooperberg, C., Stanford, J., Nevitt, M., . . . Mortality, C. (2003). Outcomes ascertainment and adjudication methods in the Women's Health Initiative. *Ann Epidemiol, 13*(9 Suppl), S122-128.

Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group. (1998). *Control Clin Trials, 19*(1), 61-109.

Emsley, R., & Liu, H. (2013). PARAMED: Stata module to perform causal mediation analysis using parametric regression models: Boston College Department of Economics. Retrieved from https://ideas.repec.org/c/boc/bocode/s457581.html

Hansen, J. D., Shimbo, D., Shaffer, J. A., Hong, S., Borda, T., Ventura, A., . . . Davidson, K. W. (2010). Finding the glass half full? Optimism is protective of 10-year incident CHD in a population-based study: the Canadian Nova Scotia Health Survey. *Int J Cardiol, 145*(3), 603-604. doi:10.1016/j.ijcard.2010.08.059

Howie, B., Fuchsberger, C., Stephens, M., Marchini, J., & Abecasis, G. R. (2012). Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet, 44*(8), 955-959. doi:10.1038/ng.2354

International Schizophrenia, C., Purcell, S. M., Wray, N. R., Stone, J. L., Visscher, P. M., O'Donovan, M. C., . . . Sklar, P. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature, 460*(7256), 748-752. doi:10.1038/nature08185

Izawa, S., Eto, Y., Yamada, K. C., Nakano, M., Yamada, H., Nagayama, M., . . . Nomura, S. (2011). Cynical hostility, anger expression style, and acute myocardial infarction in middle-aged Japanese men. *Behav Med, 37*(3), 81-86. doi:10.1080/08964289.2011.585547

Kubzansky, L. D., Sparrow, D., Vokonas, P., & Kawachi, I. (2001). Is the glass half empty or half full? A prospective study of optimism and coronary heart disease in the normative aging study. *Psychosom Med, 63*(6), 910-916.

Langer, R. D., White, E., Lewis, C. E., Kotchen, J. M., Hendrix, S. L., & Trevisan, M. (2003). The Women's Health Initiative Observational Study: baseline characteristics of participants and reliability of baseline measures. *Ann Epidemiol, 13*(9 Suppl), S107-121.

Mahasneh, A. M. A.-Z., Z. H.; Batayeneh, O. T. (2013). The Relationship between Optimism-Pessimism and Personality Traits among Students in the Hashemite University. *International Education Studies, 6*(8).

Malouff, J. M. S., N. S. (2017). Can psychological interventions increase optimism? A meta-analysis. *The Journal of Positive Psychology, 12*(6), 594-604.

Markovitz, J. H. (1998). Hostility is associated with increased platelet activation in coronary heart disease. *Psychosom Med, 60*(5), 586-591.

McCrae, R. R., & Costa, P. T., Jr. (1987). Validation of the five-factor model of personality across instruments and observers. *J Pers Soc Psychol, 52*(1), 81-90.

Merjonen, P., Keltikangas-Jarvinen, L., Jokela, M., Seppala, I., Lyytikainen, L. P., Pulkki-Raback, L., . . . Lehtimaki, T. (2011). Hostility in adolescents and adults: a genome-wide association study of the Young Finns. *Transl Psychiatry, 1*, e11. doi:10.1038/tp.2011.13

Mick, E., McGough, J., Deutsch, C. K., Frazier, J. A., Kennedy, D., & Goldberg, R. J. (2014). Genome-wide association study of proneness to anger. *PLoS One, 9*(1), e87257. doi:10.1371/journal.pone.0087257

Mozaffarian, D., Benjamin, E. J., Go, A. S., Arnett, D. K., Blaha, M. J., Cushman, M., . . . Stroke Statistics, S. (2015). Heart disease and stroke statistics--2015 update: a report from the American Heart Association. *Circulation, 131*(4), e29-322. doi:10.1161/CIR.0000000000000152

NCHS., C. (2015). Underlying Cause of Death 1999-2013 on CDC WONDER Online Database. Retrieved Feb 3, 2017

Nikpay, M., Goel, A., Won, H. H., Hall, L. M., Willenborg, C., Kanoni, S., . . . Farrall, M. (2015). A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet, 47*(10), 1121-1130. doi:10.1038/ng.3396

O'Donovan, A., Lin, J., Tillie, J., Dhabhar, F. S., Wolkowitz, O. M., Blackburn, E. H., & Epel, E. S. (2009). Pessimism correlates with leukocyte telomere shortness and elevated interleukin-6 in post-menopausal women. *Brain Behav Immun, 23*(4), 446-449. doi:10.1016/j.bbi.2008.11.006

Pankalainen, M. T., Kerola, T. V., & Hintikka, J. J. (2015). Pessimism and the risk for coronary heart disease among middle-aged and older Finnish men and women: a ten-year follow-up study. *BMC Cardiovasc Disord, 15*, 113. doi:10.1186/s12872-015-0097-y

Pearl, J. (2001). *Direct and Indirect Effects.* Retrieved from San Francisco, CA: http://ftp.cs.ucla.edu/pub/stat_ser/R273-U.pdf

Plomin, R. S., M.F.; Bergeman, C.S.; Pedersen, N.L.; Nesselroade, J.R.; McClearn, G.E. (1992). Optimism, pessimism and mental health: A twin/adoption analysis. . *Personality and Individual Differences, 13*(8).

Raikkonen, K., Matthews, K. A., Flory, J. D., Owens, J. F., & Gump, B. B. (1999). Effects of optimism, pessimism, and trait anxiety on ambulatory blood pressure and mood during everyday life. *J Pers Soc Psychol, 76*(1), 104-113.

Robins, J. M., & Greenland, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology, 3*(2), 143-155.

Roy, B., Diez-Roux, A. V., Seeman, T., Ranjit, N., Shea, S., & Cushman, M. (2010). Association of optimism and pessimism with inflammation and hemostasis in the Multi-Ethnic Study of Atherosclerosis (MESA). *Psychosom Med, 72*(2), 134-140. doi:10.1097/PSY.0b013e3181cb981b

Scheier, M. F., Carver, C. S., & Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a reevaluation of the Life Orientation Test. *J Pers Soc Psychol, 67*(6), 1063-1078.

Scheier, M. F., Weintraub, J. K., & Carver, C. S. (1986). Coping with stress: divergent strategies of optimists and pessimists. *J Pers Soc Psychol, 51*(6), 1257-1264.

Scherwitz, L. W., Perkins, L. L., Chesney, M. A., Hughes, G. H., Sidney, S., & Manolio, T. A. (1992). Hostility and health behaviors in young adults: the CARDIA Study. Coronary Artery Risk Development in Young Adults Study. *Am J Epidemiol, 136*(2), 136-145.

Schunkert, H., Konig, I. R., Kathiresan, S., Reilly, M. P., Assimes, T. L., Holm, H., . . . Samani, N. J. (2011). Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet, 43*(4), 333-338. doi:10.1038/ng.784

Sharot, T., Riccardi, A. M., Raio, C. M., & Phelps, E. A. (2007). Neural mechanisms mediating optimism bias. *Nature, 450*(7166), 102-105. doi:10.1038/nature06280

Shin, C., Kim, J., Yi, H., Lee, H., Lee, J., & Shin, K. (2005). Relationship between trait-anger and sleep disturbances in middle-aged men and women. *J Psychosom Res, 58*(2), 183-189. doi:10.1016/j.jpsychores.2004.07.002

Siegler, I. C., Peterson, B. L., Barefoot, J. C., & Williams, R. B. (1992). Hostility during late adolescence predicts coronary risk factors at mid-life. *Am J Epidemiol, 136*(2), 146-154.

Smith, G. D., & Ebrahim, S. (2003). 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol, 32*(1), 1-22.

Solovieff, N., Cotsapas, C., Lee, P. H., Purcell, S. M., & Smoller, J. W. (2013). Pleiotropy in complex traits: challenges and strategies. *Nat Rev Genet, 14*(7), 483-495. doi:10.1038/nrg3461

Steptoe, A., Cropley, M., Griffith, J., & Kirschbaum, C. (2000). Job strain and anger expression predict early morning elevations in salivary cortisol. *Psychosom Med, 62*(2), 286-292.

Steptoe, A., Wright, C., Kunz-Ebrecht, S. R., & Iliffe, S. (2006). Dispositional optimism and health behaviour in community-dwelling older people: associations with healthy ageing. *Br J Health Psychol, 11*(Pt 1), 71-84. doi:10.1348/135910705X42850

Stewart, J. C., Janicki-Deverts, D., Muldoon, M. F., & Kamarck, T. W. (2008). Depressive symptoms moderate the influence of hostility on serum interleukin-6 and C-reactive protein. *Psychosom Med, 70*(2), 197-204. doi:10.1097/PSY.0b013e3181642a0b

Thomas, K. S., Nelesen, R. A., & Dimsdale, J. E. (2004). Relationships between hostility, anger expression, and blood pressure dipping in an ethnically diverse sample. *Psychosom Med, 66*(3), 298-304.

Tindle, H. A., Chang, Y. F., Kuller, L. H., Manson, J. E., Robinson, J. G., Rosal, M. C., . . . Matthews, K. A. (2009). Optimism, cynical hostility, and incident coronary heart disease and mortality in the Women's Health Initiative. *Circulation, 120*(8), 656-662. doi:10.1161/CIRCULATIONAHA.108.827642

Tinker, L. F., Rosal, M. C., Young, A. F., Perri, M. G., Patterson, R. E., Van Horn, L., . . . Wu, L. (2007). Predictors of dietary change and maintenance in the Women's Health Initiative Dietary Modification Trial. *J Am Diet Assoc, 107*(7), 1155-1166. doi:10.1016/j.jada.2007.04.010

Valeri, L., & Vanderweele, T. J. (2013). Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros. *Psychol Methods, 18*(2), 137-150. doi:10.1037/a0031034

van den Berg, S. M., de Moor, M. H., Verweij, K. J., Krueger, R. F., Luciano, M., Arias Vasquez, A., . . . Boomsma, D. I. (2016). Meta-analysis of Genome-Wide Association Studies for Extraversion: Findings from the Genetics of Personality Consortium. *Behav Genet, 46*(2), 170-182. doi:10.1007/s10519-015-9735-5

Vanderweele, T. J. (2015). *Explanation in causal inference : methods for mediation and interaction.* . New York: Oxford University Press.

VanderWeele, T. J. (2016). Mediation Analysis: A Practitioner's Guide. *Annu Rev Public Health, 37*, 17-32. doi:10.1146/annurev-publhealth-032315-021402

Vella, E. J., & Friedman, B. H. (2007). Autonomic characteristics of defensive hostility: reactivity and recovery to active and passive stressors. *Int J Psychophysiol, 66*(2), 95-101. doi:10.1016/j.ijpsycho.2007.03.014

Wong, J. M., Na, B., Regan, M. C., & Whooley, M. A. (2013). Hostility, health behaviors, and risk of recurrent events in patients with stable coronary heart disease: findings from the Heart and Soul Study. *J Am Heart Assoc, 2*(5), e000052. doi:10.1161/JAHA.113.000052

Zaitlen, N., Lindstrom, S., Pasaniuc, B., Cornelis, M., Genovese, G., Pollack, S., . . . Price, A. L. (2012). Informed conditioning on clinical covariates increases power in case-control association studies. *PLoS Genet, 8*(11), e1003032. doi:10.1371/journal.pgen.1003032

Table 3.1: Baseline and follow-up characteristics of WHI-AA, EA and HA samples

	African Americans (n=6,336)	**European Americans (n=13,735)**	**Hispanic Americans (n=2,935)**
Baseline Characteristics			
Age (years), *mean (SD)*	60.8 (6.7)	66.0 (6.7)	59.8 (6.5)
Education, *N (%)*			
> high school, < college	1950 (31.2)	4797 (35.1)	1091 (37.8)
any college	1690 (27.0)	5317 (38.9)	722 (25.0)
> college	1897 (30.3)	3451 (25.2)	457 (15.8)
Annual household income, *N (%)*			
20000-49999	2621 (44.0)	6621 (50.9)	1080 (40.3)
50000-99999	1612 (27.0)	3240 (24.9)	555 (20.7)
≥ 100000	336 (5.6)	782 (6.0)	141 (5.3)
Living alone, *N (%)*	1759 (28.2)	3724 (27.3)	478 (16.6)
Religious service attendance, *N (%)*			
1-3 times/month	1497 (23.8)	2689 (19.6)	675 (23.3)
≥ 1 time/ week	3771 (60.0)	6590 (48.2)	1506 (51.9)
CT status, *N (%)*	3582 (56.5)	11600 (84.5)	1517 (51.7)
Smoking, *N (%)*			
ever	2471 (39.7)	5449 (40.1)	873 (30.2)
current	673 (10.8)	976 (7.2)	181 (6.3)
Alcohol consumption, *N (%)*			
ever	3097 (49.5)	8231 (60.3)	1648 (56.9)
current	271 (4.3)	1756 (12.9)	157 (5.4)
Physical activity, *N (%)*			
middle	2928 (47.3)	6546 (51.0)	1308 (46.8)
high	1121 (18.1)	2810 (21.9)	570 (20.4)
Body mass index (kg/m^2), *median (Q1, Q3)*	29.9 (26.4, 34.6)	27.5 (24.2, 31.6)	27.8 (24.9, 31.6)
Waist circumference ≥ 88cm, *N (%)*	3511 (55.6)	6462 (47.2)	1183 (40.5)
Diabetes, *N (%)*	664 (10.5)	717 (5.2)	192 (6.6)
Hypertension, *N (%)*	2783 (43.9)	3531 (25.7)	565 (19.3)
Hypercholesterolemia, *N (%)*	813 (13.4)	1645 (13.0)	335 (12.4)
Optimism score, *mean (SD)*	23.5 (3.4)	23.4 (3.3)	22.3 (3.6)
Cynical hostility score, *mean (SD)*	4.7 (3.0)	3.5 (2.7)	4.5 (3.3)
Follow-up Characteristics			
Follow-up duration (days), *mean (SD)*	5409 (976)	5404 (1013)	5267 (1062)
Age at end of follow-up (years), mean (SD)	75.6 (6.7)	80.8 (6.5)	74.2 (6.8)
All CHD events			
number of all CHD during follow-up, *N (%)*	507 (8.0)	1664 (12.1)	192 (6.5)
age at first all CHD event, *mean (SD)*	70.6 (7.8)	75.3 (7.6)	70.1 (7.9)

Acute CHD events			
number of acute CHD during follow-up, *N (%)*	241 (3.8)	981 (7.1)	80 (2.7)
age at first acute CHD event, *mean (SD)*	72.2 (8.2)	76.2 (8.1)	72.2 (8.4)

Table 3.2: Associations between the liability of CHD and psychological attitudes

	Conditional liability of all CHD		**Conditional liability of acute CHD**	
	Beta (S.E.)	**P**	**Beta (S.E.)**	**P**
Optimism				
African Americans	0.0031 (0.0023)	0.17	0.0023 (0.0018)	0.19
European Americans	-0.0043 (0.0018)	0.02 *	-0.0034 (0.0016)	0.03 *
Hispanic Americans	-0.0003 (0.0031)	0.92	0.0003 (0.0022)	0.90
Cynical Hostility				
African Americans	-0.0019 (0.0026)	0.47	0.0007 (0.0020)	0.72
European Americans	0.0030 (0.0023)	0.19	0.0039 (0.0019)	0.04 *
Hispanic Americans	0.0006 (0.0034)	0.87	0.0008 (0.0024)	0.74

All models adjusted for age, education, living alone, annual household income, religious service attendance, diabetes, hypertension, hypercholesterolemia, smoking, alcohol consumption, physical activity, waist circumference, body mass index, OS/CT status.
Asterisk indicates significance at $P<0.05$.

Table 3.3: Associations of CHD-PRS with psychological attitudes and the liability of CHD in the WHI-EA sample

PRS for CHD	**Target Phenotypes**			
	Psychological Attitudes		**Conditional Liability of CHD**	
P-threshold	**Optimism**	**Cynical hostility**	**All CHD**	**Acute CHD**
$P_T < 0.0001$	t=-3.43 P=6.09E-04** R^2=0.0009	t=-0.28 P=0.78 R^2=5.76E-06	t=5.87 P=4.51E-09** R^2=0.0025	t=4.38 P=1.20E-05** R^2=0.0014
$P_T < 0.001$	t=-0.57 P=0.57 R^2=2.40E-05	t=0.34 P=0.73 R^2=8.41E-06	t=2.01 P=0.05* R^2=0.0003	t=0.62 P=0.53 R^2=2.81E-05
$P_T < 0.01$	t=0.47 P=0.64 R^2=1.60E-05	t=0.65 P=0.52 R^2=3.14E-05	t=1.91 P=0.06 R^2=0.0003	t=0.18 P=0.86 R^2=2.25E-06

All models adjusted for the top 10 PCs.
Single asterisk indicates significance at p<0.05.
Double asterisks indicate significance after Bonferroni correction.

Table 3.4: Total, direct and indirect effects of CHD-PRS on the liability of CHD mediated by optimism in the WHI-EA sample

Exposure	**Mediator**	**Outcome**	**Total effect (95% BCCI)**	**Direct effect (95% BCCI)**	**Indirect effect (95% BCCI)**	**PM**
CHD-PRS (75^{th} vs. 25^{th} percentiles)	Optimism	Liability of all CHD	0.0446 (0.0281, 0.0596)	0.0440 (0.0274, 0.0591)	0.0006 (0.0001, 0.0014)	1.4%
		Liability of acute CHD	0.0293 (0.0168, 0.0435)	0.0288 (0.0164, 0.0430)	0.0005 (0.0001, 0.0012)	1.7%
CHD-PRS (90^{th} vs. 10^{th} percentiles)		Liability of all CHD	0.0842 (0.0564, 0.1144)	0.0831 (0.0541, 0.1133)	0.0012 (0.0003, 0.0028)	1.4%
		Liability of acute CHD	0.0552 (0.0300, 0.0788)	0.0543 (0.0287, 0.0777)	0.0009 (0.0001, 0.0021)	1.7%

Figure 3.1: The potential causal relationships between CHD-PRS, optimism, and CHD

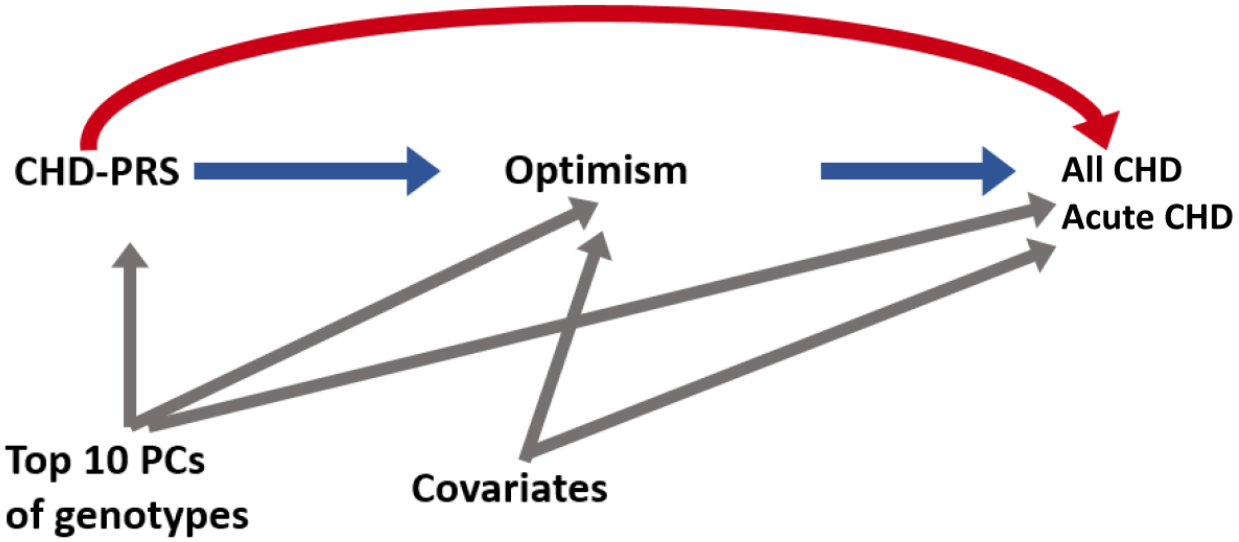
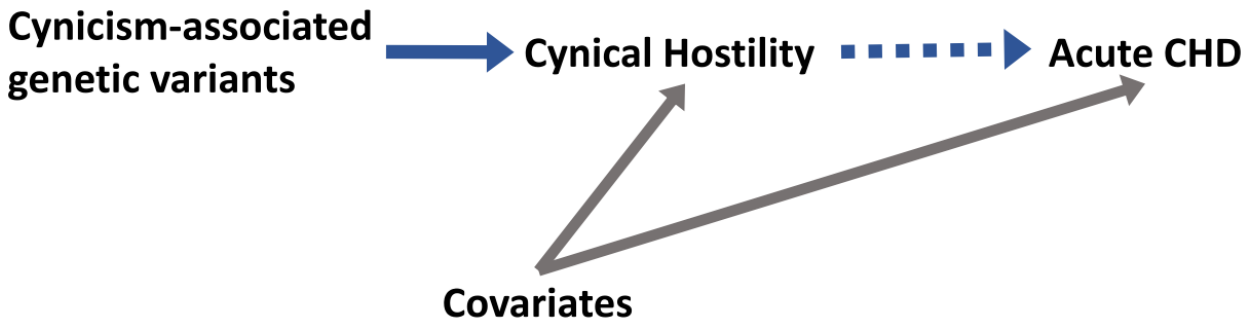


Figure 3.2: Cynicism-associated genetic variants that are not directly associated with CHD may be a potential instrumental variable for Mendelian randomization to determine causal effect of cynical hostility on acute CHD



Causal relationships between genetics, psychological attitudes and coronary heart disease

Supplementary Information

SUPPLEMENTARY METHODS

Sensitivity Analyses of Unmeasured Confounding

The counterfactual-based mediation analysis assumes no unmeasured confounding between (1) exposure and mediator, (2) exposure and outcome, and (3) mediator and outcome (T.J. Vanderweele, 2015). In our mediation analyses with genetic scores as the exposures, since we had adjusted for the top principal components (PCs) of ancestry to address possible population stratification, assumptions (1) and (2) would probably hold. However, the assumption of no unmeasured confounding between (3) mediator and outcome might be less plausible, and the effect estimates would probably be biased.

In order to evaluate the robustness of the mediation analyses to unmeasured confounding for the mediator-outcome relationship, we conducted sensitivity analyses to calculate how much direct and indirect effect estimates would be expected to change under the existence of unmeasured mediator-outcome confounding. Specifically, given a hypothetical unmeasured confounder of the mediator-outcome relationship, U, with particular correlations with the mediator and the outcome, we examined what the true direct and indirect effect estimates would be, if we were able to adjust for U.

For each mediation analysis, we generated four standard normal variables (i.e., mean=0 and variance=1), with particular correlations with the mediator and the outcome, as hypothetical confounders (Table S3.4-3.5). The first hypothetical confounder has correlation of 0.1 with the mediator and correlation of 0.1 with the outcome. The second hypothetical confounder has correlation of 0.2 with the mediator and correlation of 0.2 with the outcome. The third hypothetical confounder has correlation of -0.1 with the mediator and correlation of 0.1 with the outcome. The fourth hypothetical confounder has correlation of -0.2 with the mediator and correlation of 0.2 with the outcome. We compared the direct and indirect effect estimates before and after adjusting for each of these hypothetical confounders, to assess the potential impact of unmeasured confounding on each mediation analysis.

SUPPLEMENTARY RESULTS

Sensitivity Analyses of Unmeasured Confounding

In the mediation analysis for EA individuals, with CHD-PRS as the exposure, optimism score as the mediator, and the age-conditional liability of all CHD as the outcome reported in the main text, the estimated direct and indirect effects (95%CI) before adjusting for the hypothetical unmeasured confounder were 0.0440 (0.0274, 0.0591) and 0.0006 (0.0001, 0.0014), respectively. The effect estimates after adjusting for each hypothetical confounder U are shown in Table S3.5. Under adjustment of a hypothetical confounder with positive correlations with both mediator and outcome, the estimated proportion mediated would be greater than that without considering the hypothetical confounder. Under adjustment of a hypothetical confounder with correlations of -0.1 and 0.1 with mediator and outcome respectively, the estimated proportion mediated would be smaller but remain significantly greater than zero. However, under adjustment of a stronger hypothetical confounder with correlations of -0.2 and 0.2 with mediator and outcome respectively, the estimated proportion mediated would become negative, and this may change our conclusions.

In the mediation analysis for EA individuals, with CHD-PRS as the exposure, optimism score as the mediator, and the age-conditional liability of acute CHD as the outcome reported in the main text, the estimated direct and indirect effects (95%CI) before adjusting for the hypothetical unmeasured confounder were 0.0288 (0.0164, 0.0430) and 0.0005 (0.0001, 0.0012), respectively. The effect estimates after adjusting for hypothetical unmeasured confounders are shown in Table S3.6. Similar to the results of sensitivity analysis for the mediation analysis with the age-conditional liability of all CHD as the outcome (Table S3.5), in the presence of an unmeasured confounder associated with higher risk of CHD and greater optimism, the true mediation effect of optimism would be greater than that estimated in the original mediation analysis. However, if the associations of the unmeasured confounder with optimism and CHD risk were in different directions, our estimated proportion mediated would overestimate the true mediation.

SUPPLEMENTARY TABLES

Table S3.1: Description of the six WHI GWAS samples

	SHARe	**GARNET**	**WHIMS+**	**HIPFx**	**MOPMAP**	**GECCO**
Sample size	7470 AA 3348 HA	3727	5687	2841	2840	2083
Ethnicity	African American and Hispanic	White	White	White	White	White
GWAS platform	Affymetrix 6.0	Illumina HumanOmni1-Quad v1-0 B	HumanOmniExpressExome-8v1_B	Illumina 550K and 610K	Affymetrix Gene Titan, Axiom Genome-Wide Human CEU I Array Plate	Illumina 610 and Cytochip 370K
Design	Cohort	Case-control (4 case groups)	Cohort	Case-control	Case-Control	Case-control
Phenotype for cases	NA	Type 2 Diabetes, Myocardial Infarction, Stroke, Venous Thrombosis	NA	Hip Fracture	Ventricular Ectopy (ever)	Colorectal cancer

Table S3.2: QC and Imputation for the six WHI GWAS samples

	SHARe	**GARNET**	**WHIMS+**	**HIPFx**	**MOPMAP**	**GECCO**
Minimal sample call rate	95%	98%	97%	98%	95%	97%
Minimal SNP call rate	90%	98%	98%	98%	90%	98%
Hardy Weinberg P-value cut-off below which SNPs are excluded	1e-6	1e-4	1e-4	1e-4	1e-6	1e-4
Samples used for Hardy Weinberg calculations	All samples, separate for AA and HA	Controls	All	Controls	All	Controls
Minimum allele frequency cut-off	1%	None	1%	1%	0.5%	5%
Imputation software	MACH	minimac	minimac	minimac	minimac	minimac
Imputation reference panel	1kGP v3.20101123	1kGP v3.20101123	1kGP v3.20101123	1kGP v3.20101123	1kGP v3.20101123	1kGP v2.20101123

Table S3.3: Associations of CHD-PRS with psychological attitudes and the liability of CHD in the AA subsample

PRS for CHD		**Target Phenotypes**			
		Psychological Attitudes		**Conditional Liability of CHD**	
P-threshold for EUR-PRS	**P-threshold for AA-PRS**	**Optimism**	**Cynical hostility**	**All CHD**	**Acute CHD**
$P_T < 0.0001$	$P_T < 0.0001$	t=0.36 P=0.73 R^2=5.22E-06	t=0.29 P=0.77 R^2=4.76E-06	t=1.67 P=0.16 R^2=8.23E-05	t=1.35 P=0.29 R^2=5.63E-05
	$P_T < 0.001$	t=-0.21 P=0.82 R^2=3.40E-06	t=0.30 P=0.76 R^2=7.58E-06	t=1.86 P=0.07 R^2=0.0002	t=1.52 P=0.20 R^2=7.81E-05
	$P_T < 0.01$	t=-0.46 P=0.65 R^2=1.53E-05	t=0.54 P=0.59 R^2=2.41E-05	t=3.31 P=7.20E-04** R^2=0.0011	t=2.98 P=1.53E-03** R^2=0.0009
$P_T < 0.001$	$P_T < 0.0001$	t=-0.39 P=0.70 R^2=6.38E-06	t=0.35 P=0.74 R^2=5.16E-06	t=0.24 P=0.81 R^2=4.79E-06	t=0.29 P=0.78 R^2=6.01E-06
	$P_T < 0.001$	t=0.09 P=0.97 R^2=2.07E-06	t=0.38 P=0.72 R^2=6.21E-06	t=0.58 P=0.55 R^2=2.21E-05	t=0.44 P=0.68 R^2=1.33E-05
	$P_T < 0.01$	t=-0.47 P=0.64 R^2=1.60E-05	t=0.65 P=0.52 R^2=3.14E-05	t=1.92 P=0.06 R^2=0.0003	t=0.21 P=0.83 R^2=2.54E-06
$P_T < 0.01$	$P_T < 0.0001$	t=-0.43 P=0.65 R^2=7.04E-06	t=-0.18 P=0.87 R^2=3.06E-06	t=0.85 P=0.37 R^2=5.36E-05	t=0.32 P=0.74 R^2=8.67E-06
	$P_T < 0.001$	t=-0.27 P=0.78 R^2=5.40E-06	t=0.34 P=0.73 R^2=8.32E-06	t=0.27 P=0.79 R^2=0.0003	t=0.57 P=0.55 R^2=2.81E-05
	$P_T < 0.01$	t=-0.24 P=0.81 R^2=4.60E-06	t=0.55 P=0.61 R^2=1.24E-05	t=0.43 P=0.68 R^2=9.23E-06	t=0.54 P=0.57 R^2=2.25E-05

EUR-PRS: We used CARDIoGRAM GWAS results as the training dataset to calculate PRS with European training data with different p-value thresholds (0.0001, 0.001, and 0.01) in the WHI-AA target sample.
AA-PRS: We used a five-fold cross-validation procedure to calculate PRS with a p-value threshold of 0.0001, 0.001, or 0.01 in the WHI-AA subsample.
CHD-PRS, generated by using MultiPRS: We performed a 2-dimensional grid search to optimize the linear combination via an in-sample fit on the WHI-AA validation sample, with different P-value thresholds for EUR-PRS and AA-PRS. We examined the prediction accuracy (adjusted R^2) for the liability of CHD of the 9 (3x3) combinations for the WHI-AA sample, and used the fitted beta coefficients as the weights to generate the final CHD-PRS for each AA participant. We also tested the association of each CHD-PRS with each psychological attitude.

Table S3.4: Associations of CHD-PRS with psychological attitudes and the liability of CHD in the HA subsample

PRS for CHD		**Target Phenotypes**			
		Psychological Attitudes		**Conditional Liability of CHD**	
P-threshold for EUR-PRS	**P-threshold for HA-PRS**	**Optimism**	**Cynical hostility**	**All CHD**	**Acute CHD**
$P_T < 0.0001$	$P_T < 0.0001$	t=-0.35 P=0.73 R^2=5.12E-06	t=0.12 P=0.91 R^2=2.36E-06	t=1.07 P=0.36 R^2=5.63E-05	t=0.88 P=0.46 R^2=3.32E-05
	$P_T < 0.001$	t=-0.41 P=0.67 R^2=8.40E-06	t=0.20 P=0.86 R^2=3.58E-06	t=2.47 P=0.01* R^2=0.0007	t=2.01 P=0.04* R^2=0.0004
	$P_T < 0.01$	t=-0.46 P=0.65 R^2=1.53E-05	t=0.54 P=0.59 R^2=2.41E-05	t=1.89 P=0.05 R^2=0.0003	t=1.42 P=0.26 R^2=7.45E-05
$P_T < 0.001$	$P_T < 0.0001$	t=0.15 P=0.90 R^2=2.13E-06	t=0.32 P=0.76 R^2=4.86E-06	t=0.43 P=0.69 R^2=9.36E-06	t=0.46 P=0.66 R^2=1.41E-05
	$P_T < 0.001$	t=-0.38 P=0.70 R^2=4.96E-06	t=0.04 P=0.98 R^2=1.20E-06	t=0.52 P=0.57 R^2=2.19E-05	t=0.45 P=0.66 R^2=1.37E-05
	$P_T < 0.01$	t=-0.57 P=0.52 R^2=3.21E-05	t=0.65 P=0.50 R^2=3.64E-05	t=0.56 P=0.53 R^2=2.46E-05	t=0.20 P=0.83 R^2=3.01E-06
$P_T < 0.01$	$P_T < 0.0001$	t=-0.27 P=0.77 R^2=5.34E-06	t=0.08 P=0.96 R^2=1.66E-06	t=1.15 P=0.28 R^2=7.06E-05	t=0.87 P=0.47 R^2=3.28E-05
	$P_T < 0.001$	t=-0.07 P=0.96 R^2=1.49E-06	t=0.30 P=0.75 R^2=7.63E-06	t=0.93 P=0.39 R^2=4.84E-05	t=0.67 P=0.53 R^2=2.79E-05
	$P_T < 0.01$	t=-0.34 P=0.72 R^2=4.49E-06	t=0.43 P=0.68 R^2=7.04E-06	t=0.40 P=0.70 R^2=8.14E-06	t=0.56 P=0.53 R^2=2.45E-05

EUR-PRS: We used CARDIoGRAM GWAS results as the training dataset to calculate PRS with European training data with different p-value thresholds (0.0001, 0.001, and 0.01) in the WHI-HA target sample.
HA-PRS: We used a five-fold cross-validation procedure to calculate PRS with a p-value threshold of 0.0001, 0.001, or 0.01 in the WHI-HA sample.
CHD-PRS, generated by using MultiPRS: We performed a 2-dimensional grid search to optimize the linear combination via an in-sample fit on the WHI-HA validation sample, with different P-value thresholds for EUR-PRS and AA-PRS. We examined the prediction accuracy (adjusted R^2) for the liability of CHD of the 9 (3x3) combinations for the WHI-HA sample, and used the fitted beta coefficients as the weights to generate the final CHD-PRS for each HA participant. We also tested the association of each CHD-PRS with each psychological attitude.

Table S3.5: The estimated direct and indirect effect beta (95% CI) of the relationship between the CHD-PRS, optimism, and the conditional liability of all CHD after adjusting for a hypothetical confounder U

Unmeasured Confounder		**Mediation Analyses after additionally adjusting for the hypothetical unmeasured confounder**			
rUM	**rUY**	**Total effect (95% BCCI)**	**Direct effect (95% BCCI)**	**Indirect effect (95% BCCI)**	**PM**
0.1	0.1	0.0438 (0.0282, 0.0580)	0.0430 (0.0274, 0.0569)	0.0008 (0.0003, 0.0018)	1.9%
0.2	0.2	0.0425 (0.0275, 0.0576)	0.0409 (0.0259, 0.0560)	0.0016 (0.0006, 0.0028)	3.7%
-0.1	0.1	0.0435 (0.0284, 0.0593)	0.0431 (0.0277, 0.0583)	0.0004 (0.0001, 0.0011)	1.0%
-0.2	0.2	0.0411 (0.0257, 0.0570)	0.0413 (0.0258, 0.0572)	-0.0002 (-0.0007, 0.0001)	-0.5%

rUM: the Pearson's correlation coefficient between the hypothetical confounder and the mediator
rUY: the Pearson's correlation coefficient between the hypothetical confounder and the outcome
The estimated direct and indirect effects (95%CI) before adjusting for the hypothetical unmeasured confounder were 0.0440 (0.0274, 0.0591) and 0.0006 (0.0001, 0.0014), respectively.
The estimated proportion mediated before adjusting for the hypothetical unmeasured confounder was 1.4%.

Table S3.6: The estimated direct and indirect effect beta (95% CI) of the relationship between the CHD-PRS, optimism, and the conditional liability of acute CHD after adjusting for a hypothetical confounder U

Unmeasured Confounder		**Mediation Analyses after additionally adjusting for the hypothetical unmeasured confounder**			
rUM	**rUY**	**Total effect (95% BCCI)**	**Direct effect (95% BCCI)**	**Indirect effect (95% BCCI)**	**PM**
0.1	0.1	0.0287 (0.0155, 0.0425)	0.0280 (0.0149, 0.0423)	0.0007 (0.0002, 0.0015)	2.3%
0.2	0.2	0.0278 (0.0150, 0.0411)	0.0266 (0.0134, 0.0396)	0.0013 (0.0005, 0.0023)	4.5%
-0.1	0.1	0.0283 (0.0163, 0.0425)	0.0281 (0.0161, 0.0422)	0.0003 (8.0E-06, 0.0008)	1.0%
-0.2	0.2	0.0266 (0.0145, 0.0405)	0.0268 (0.0147, 0.0405)	-0.0002 (-0.0007, 0.00003)	-0.8%

rUM: the Pearson's correlation coefficient between the hypothetical confounder and the mediator
rUY: the Pearson's correlation coefficient between the hypothetical confounder and the outcome
The estimated direct and indirect effects (95%CI) before adjusting for the hypothetical unmeasured confounder were 0.0288 (0.0164, 0.0430) and 0.0005 (0.0001, 0.0012), respectively.
The estimated proportion mediated before adjusting for the hypothetical unmeasured confounder was 1.7%.

REFERENCES

Vanderweele, T. J. (2015). *Explanation in causal inference : methods for mediation and interaction.* . New York: Oxford University Press.